

68015

From: Chan, Christina
Sent: Tuesday, June 04, 2002 1:46 PM
To: Sorbello, Eleanor; STIC-Biotech/ChemLib
Subject: RE: Rush seq search 09/515,369B

Please rush. Thanks Chris

-----Original Message-----

From: Sorbello, Eleanor
Sent: Tuesday, June 04, 2002 1:28 PM
To: Chan, Christina
Cc: Sorbello, Eleanor
Subject: Rush seq search 09/515,369B

Christina,
Can you please request a rush seq. search of (1) nucleotide SEQ. ID. NO: 1 and (2) any matches from T at position -2241 to the C at posn. 0 of Seq. Id NO: 1

Eleanor Sorbello
Art Unit 1632 ✓
CM1 12D14 ✓
(703) 308-6043

12E12 ✓
Numbering does not correspond to our database system
of numbering.

In our files the first 10 positions are: taatacga₁ct₁₀.
There is no zero position and there are no negative numbers.

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

TYPE OF SEARCH:		VENDOR/COST(where applic.)
Searcher: _____	NA Sequences: _____	STN: _____
Phone: _____	AA Sequences: _____	DIALOG: _____
Location: _____	Structures: _____	Questel/Orbit: _____
Date Picked Up: <u>6/3</u>	Bibliographic: _____	DRLink: _____
Date Completed: <u>6/6</u>	Litigation: _____	Lexis/Nexis: _____
Searcher Prep/Review: _____	Full text: _____	Sequence Sys.: _____
Clerical: _____	Patent Family: _____	WWW/Internet: _____
Online time: _____	Other: _____	Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 6, 2002, 08:23:33 ; Search time 4143.56 Seconds
(without alignments)
11345.160 Million cell updates/sec

Title: US-09-515-369B-1
Perfect score: 2286
Sequence: 1 taatacagactacactataggg.....tgacttcacggcgtgggacg 2286

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues 3595312
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.cm.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	2255	98.6	2256	9	AF217405	
2	2208.6	96.6	193317	2	AC098935	
c	3	2208.6	96.6	229302	2	AC023534
4	2151.4	94.1	180048	2	AC068122	
5	1872.2	81.9	11114	9	AY062931	
6	1365.2	59.7	7025	9	AF276916	
7	91.8	4.0	29034	9	HS243947	
c	8	91.8	4.0	174210	2	AC108684
c	9	91.2	4.0	137769	9	AC002090
10	90.8	4.0	193126	9	AL139275	
11	90	3.9	177957	2	AC023298	
c	12	90	3.9	178477	2	AC022801
13	89.2	3.9	192695	2	AC098645	
14	89.2	3.9	181009	2	AC011060	
15	89.2	3.9	183483	9	AL161628	
16	89.2	3.9	194590	2	HS315114	
17	88.4	3.9	168486	9	CNS0578N	
18	88.2	3.9	149973	2	AC023992	
19	88.2	3.9	158905	9	AC005666	
c	20	88.2	3.9	163242	2	AC068438
c	21	87.4	3.8	181134	2	AC074343
22	87	3.8	62497	2	AC103865	
c	23	87	3.8	178756	2	AC104420
c	24	86	3.8	116516	2	AC109809
c	25	85.4	3.7	86025	9	AL158175
c	26	85.4	3.7	119361	2	AC092771
c	27	85.4	3.7	142192	2	AC092713
28	85.4	3.7	158341	2	AC021733	
29	85.4	3.7	160671	9	AC090451	
c	30	85.4	3.7	162625	9	AC006083
c	31	85	3.7	127936	2	AC022099
c	32	85	3.7	143536	2	AC020846
c	33	85	3.7	172830	9	AL137072
c	34	84.6	3.7	59319	2	AC103842
c	35	84.6	3.7	114467	9	AP000462
c	36	84.6	3.7	154236	2	AC051650
37	84.6	3.7	157322	9	AC012103	
c	38	84.6	3.7	157399	2	AP003167
c	39	84.6	3.7	170965	9	AB017654
c	40	84.4	3.7	144714	9	AC020891
c	41	84.4	3.7	145886	9	AP003438
c	42	84.4	3.7	175830	2	AC090547
c	43	84.4	3.7	188764	2	AC021238
c	44	84	3.7	170893	2	AC027537
c	45	83.8	3.7	37321	9	AC005498

ALIGNMENTS

RESULT 1

AF217405

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

AF217405 Homo sapiens suppression of tumorigenicity 16 protein (ST16) gene, promoter and partial sequence.

AF217405

AF217405.1 GI:10185703

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2256)

Madireddi, M.T., Dent, P. and Fisher, P.B.

AP-1 and C/EBP transcription factors contribute to mda-7 gene promoter activity during human melanoma differentiation

J. Cell. Physiol. 185 (1), 36-46 (2000)

20400556

10942517

2 (bases 1 to 2256)

Madireddi, M.T. and Fisher, P.B.

Direct Submission

Submitted (17-DEC-1999) Urology, Columbia University, 630 West,

168th Street, BB 1505, New York, NY 10032, USA

FEATURES
source

Location/Qualifiers
1..2256

/organism="Homo sapiens"

/db_xref="taxon:9606"

/tissue.type="placenta"

1..2208

/gene="ST16"

1..>2256

/gene="ST16"

/note="mda-7"

2209..>2256

/gene="ST16"

BASE COUNT 539 a 584 c 503 g 630 t
ORIGIN

Query Match 98.6%; Score 2255; DB 9; Length 2256;
Best Local Similarity 100.0%; Fred. No. 0;
Matches 2255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 tcacctttgaacccagggtgctgctgccctccaaagcttgtaactcataactagattctcaa 91
DB 1 TCACCTTTTGAACCCAGGCTGCTGCCCTCCAAAGCTTGTACTCATTAAGTACTCAA 60
QY 92 ctgattgtgggccaaggttctctaggttctctctgacattctctctgaagtaataatgc 151
DB 61 CTGATGTTGGGCCAAGGTTCTCTAGGTCTCTCTCTGACCTTCTCTGAAAGTAATAATGC 120
QY 152 tatgataagctcagcaggtgagcctgagccagcagcaactgtttgcctgaactatccatgtta 211
DB 121 TATGATAAGCTCATCGGAGGCTGAGGCCAGGCACATGTTGCTCGAATATCCATGTTA 180
QY 212 tatgattctctcagacagagtgagctactacagatcccagggtgtaacctgagccag 271
DB 181 TATGATTCTCTCTCAGACAGAGTGAGCTACTCACGATCCAGGTGTACCTTGAGGCCAG 240
QY 272 ccaagtgatccatgacatcctgctctgttccagcctgcccctttaacagctcatccca 331
DB 241 CCAAGGTGATATCCATGACCTCATGCTCTGTTCCAGCCTGCCCTTTAAACAGCTCATCCCA 300
QY 332 cctgcctgcctcccccgcctctatctgagacagtagtctagagatttcagctgcctcgggg 391
DB 301 CTGCTGCTCCCTCCCGCCTATCTGCAGACAGTAGTCTAGGATTTACAGTGCCTTGGGG 360
QY 392 ctcaatttccctcagctcctgctgtttagctgtctctgcctcccaactcaactattact 451
DB 361 CTCATTTTCCCTCTCAGCTTCCGTGCTTTAGCTGTCTCTGCTGCCACTCACCTATTACT 420
QY 452 ccagcactctcactggtctcttttctgtctcactgctctcactgctcttgacatctttatctc 511
DB 421 CCAGCACTCTCAGCTGGTCTCTTTCTGTCTCATCACTGCTCTTGACATCTTTATCTC 480
QY 512 atagtagttagttaggggtcttggtaatgccctaaatccacatggtgggaaggggag 571
DB 481 ATAGTAGTTAGTTAGGGGTCTTGTAATGCCCTAAATCCACATGTTGGGAAGGGGGAG 540
QY 572 tgggggaagagagtgctgtgtgggctgtgcttactctctggagggttaagactcgggacct 631
DB 541 TGGGGGAAGAGAGTGGCTGTGGGGCTGTGCCCTACTTCTGGAGGGTAAAGACTCGGGCCCT 600
QY 632 ccaggaacaaagattcaagctgggtggcagctatagccaagcagactgctggccaaggat 691
DB 601 CCAGGAACAAAGATTTCAGGCTGGTGGCAGCTATAGCCAAAGCAGACTGCTGGCCAGGGAT 660
QY 692 tgcgaaggagtagttttgttctgaagaaaaaaacacactgagtagagtagagga 751
DB 661 TGCAAAAGGAGTATTTGTTGCTTAAGAAATAAACACACTGAGTATGAGATGGAGGGA 720
QY 752 ggggggtgttgggtccagagagagattgggaagagctctgccaagggtgtgttctactactct 811
DB 721 GGGGGTGTGGTGGCAGAGAGATTGGGAAGAGTCTGCCAAGGGTGTGTCTACTCACTCT 780

QY 812 cctctttttttcattctccactgagctggagggcaggtatctctgttcccccaagctcattcc 871
DB 781 CCTCTTTTCTTTTCATCTCCACTGAGCTGGAGGCAGTATCTCTGTCCCCCAGCTCAGATTC 840
QY 872 ctactcccgcttcccatgctgctggagcccggttgggcaaaactctctctgtaaagaaccaga 931
DB 841 CTACTCCCGTTTCCCATGCTGGAGCCAGGTTGGGCAAACTCTCTCTGTAAAGAACCCAGA 900
QY 932 caggaaactatttagctctctgtgccatagtctcagtcacaaactactctctctccc 991
DB 901 CAGGAACATATTTTAGCTCTGTGTGCCATATGCTCTAGTCAAACTACTCATCTCTCTGCC 960
QY 992 tctgtagcacgaagaacaattagcaacaataatgtcaacaacaataatgtgacccccatgaaaa 1051
DB 961 TCTGTAGCACGAAGACAATTAGCAACAATAATGTCAACAACAATAATGTGAGCCCATGAAAA 1020
QY 1052 ctttatttattatggaatacggaaacctgaataataatgtctcttcttcttcttcttctccca 1111
DB 1021 CTTTATTTTATTTATGATACGGAACCTGAAAAATAATATGCTTTCTTTTGTATTTTCCCA 1080
QY 1112 atcattaaaaaacgttaaaactactctttaggtcgcaaggttaagccattctcagcttagc 1171
DB 1081 ATCATTAATAAACGTAATAAATACTTCTTTAGGTGCGCAAGGTTAAGCCATTTCTCAGCTTAGC 1140
QY 1172 agtggcaggtggatttggcttggacctacagttggccaatccctgattccccaaaatgt 1231
DB 1141 AGTGGCAGGCTGGATTTGGCTTGTGACCTACAGTTGGCAATCCCTGATTTCCCAAAATGT 1200
QY 1232 attcctcagggatgtgggcaataacttattgggaagtgctggattaaacagaggttaagag 1291
DB 1201 ATTCTCTAGGGATGTGGGCAATAACTTATGGGAAGTGTGGATTTAAACAGAGTTAAGAG 1260
QY 1292 catcagacatttccagagcgggctagcacatgcccagggctctcctaaactgaactcattga 1351
DB 1261 CATCAGACATTTCCAGAGGGGCTAGCACATGCCAGGGCTCTCTAACTGAGCTCATTTGGA 1320
QY 1352 ttcatctgtttcattgagggattcttgcaagacaagaattcctcacaacctagagttctgagga 1411
DB 1321 TTCACTCTTTTCATGGAGGATCTTTCGAAGACAAGAATAATTCCTCAAACTTAGAGTCTGAGGA 1380
QY 1412 ctgtgcttgggaacactgctctgcttgatgacctcactcactcagggcacatggtagaatctag 1471
DB 1381 CTGTGCTTTGGAAACACTGCTCTGCTGTATGCCCCCTCACTGGGCACATGGTAGAATCTAG 1440
QY 1472 agctgagtgccctgtagctgagataggttcagagctcttgactgacctggcctggcagcttg 1531
DB 1441 AGCTGAGTGCCTTGTCTAGCTGGAGATAGGCTCAGAGCTCTTTGACTGCCCTGGCAGTCTTG 1500
QY 1532 acacatacagctgtctgtgtcccttgagtggttcagagccacacagggccaagactagccc 1591
DB 1501 ACACATCACGCTGTCTGTGCCCTGAGTGGTTTTCAGAGCCACACAGGCCAAGACTAGGCC 1560
QY 1592 accagagcaccagcctccagcttctggtcttgacctcactcactcagctacattctcttctct 1651
DB 1561 ACCAGAGCACCAGGCTCTCCAGCTTTCTGGGCTTGTCCATGGGTACATTTCTTATTTCTT 1620
QY 1652 cctggttcccaagaacctaaagagagggcacattttgggttgagtgatataaaaccttagggac 1711
DB 1621 CCTGTTTCCAGAACCTTAAGGAGAGGCACATTTTGGTTGAGTGTATTAACCTTAGGGAC 1680
QY 1712 catgggtagctgcatgtcagggaacactcctcactctctctggtccctgatgatgaaggga 1771
DB 1681 CATGGTAGCTGTCATGTTCAGGAACACTCTCTCAACTCTCTGCTGGCTGATGGATTTAAAGGA 1740
QY 1772 gaggtactacaggttatttctcgtgtgactactgtccacagcatgaatagggcatca 1831
DB 1741 GAGGTACTTACAGGTTATTTCTCGCTGTGGACTACTGTCCACGATGAATAGGGCATCA 1800
QY 1832 ttattgaattattttgacagaagagagactgggtgtatgtgtgcacagtaataatgtattta 1891
DB 1801 TTAATTGAATTTATTTTGACAGGAAGAGACTGGTGTATGCTGTCACAGTAATAATCTATTTA 1860

QY 685 caggattcgaaggagatttttggcttaagaaataaaacaactgagatgagat 744
|||||
Db 104154 CAGGGATTGCAAGGAGTATTTGTTGCTTAAGAAATAAACAACACTGAGTATGACAT 104213
|||||
QY 745 ggaggagggggtgtgtgagagagagattgggaagagctgccaagggtgtgtctac 804
|||||
Db 104214 GGAGGGAGGGGTGTGTGTGCCAGAGAGATTGGGAAGAGTCTGCCAAGGGGTGTCTTAC 104273
|||||
QY 805 tcaactctctctttcttctcattctccactgagctgagcgaggtatcctgtcccccagct 864
|||||
Db 104274 TCACCTCTCCTCTTTCTTTTCATCTCCACTGAGCTGAGGAGCAGTATTCCTGTCCCCACGT 104333
|||||
QY 865 cacattctcaactccgtttcccatccctgaccceaggttgggcaaaactctctctgtaag 924
|||||
Db 104334 CACATTCTCTACCTCCCGTTTCCCATGCTCCGAGCCAGGTGGGCAAACTCTTCTCTGTAAG 104393
|||||
QY 925 aaccagacaggaactattttagctctgtgtgccaataagttctcagtcacaaactactcat 984
|||||
Db 104394 AACCAGACAGGAACATATTTTAGGCTCTGTGTGCCATATGGTCTCAGTCACAACTACTCAT 104453
|||||
QY 985 ctctgctctgtagcagcaagaactatagcaacaataatgtcaacaaaacatatgtgacccc 1044
|||||
Db 104454 CTCCTGCTCTGTAGCAGCAAAACAATATAGCAACAATATGTCAACAACATATGTGACCCC 104513
|||||
QY 1045 atgaaaaactttatttattatagatacgaacaaactgaataatgtcttttttgatttt 1104
|||||
Db 104514 ATGAAAACTTTATTTATATGATGATACGGAACCTGAAATAATGTCTTTTGTGATTTT 104573
|||||
QY 1105 ttccccaatcattaaaaaacgttaaaactactcttagtcgcaagtttaagccattctca 1164
|||||
Db 104574 TTCCCCAATCATTAATAAACGTAATAAACTACTCTTAGTGTGCAAGGTTAAGCCATTCTCA 104633
|||||
QY 1165 gcttgagctggcaggtcggatttggctgtgacacacagttggccaactccctgattccc 1224
|||||
Db 104634 GCTTAGCAGTGCAGGCTGGATTTGGCTTGTGACCTACAGTTGGCCAATCCTGATTCCC 104693
|||||
QY 1225 aaaaatctctcctcaggaatgggcaataacttatggaagctgagtaaaacagagt 1284
|||||
Db 104694 AAAATGATTTCTCAGGATGGGCAAAATACTTTATGGGAATGCTGGATTAAACAGAGT 104753
|||||
QY 1285 taagaagcatcagacatttccaggagcggctagcacatgccagggtctctaaactgaact 1344
|||||
Db 104754 TAAGAAGCATCAGACATTTCCAGGACGGCTAGCACATGCCAGGGCTCTCTAACTGACCT 104813
|||||
QY 1345 cact-ggattcaatctgtttcagagagattcttgaagcaagaactctcctcaaacctagag 1403
|||||
Db 104814 CATTAGGATTTCATCTGTTTCATGAGGAGTCTTGCAAGACAAGAAATTCCTCAAACTTAGAG 104873
|||||
QY 1404 tctgaggactgtgttgggaaacactctctgtgtgatgccctcactgggcacatgta 1463
|||||
Db 104874 TCTGAGGACTGTCTTTGGGAACACTGCTCTGTTGATGCCCTCAGTGGGCACATGGTA 104933
|||||
QY 1464 gaactagagctgagtcgcttgtagctggagatagggatcagagctcttgactgccctgg 1523
|||||
Db 104934 GAATCTAGAGCTGAGTGCCTTGTCTAGCTGGAGATAGGGTTCAGAGCTCTTGACTGCCCTGG 104993
|||||
QY 1524 cagtttgacacatacagctgtctgtgtccctcagtgagtggttcagagccacacagggcaag 1583
|||||
Db 104994 CAGTCTTGACACATCACGCTGTCTGTGTCCTCCTGAGTGGTTTCAGAGGCCACACAGGCCAAG 105053
|||||
QY 1584 actagccacacagagcacacagcctcccagcttctctggtgtgttccatgcgtacattccc 1643
|||||
Db 105054 ACTAGCCACCACAGACACAGCCCTCCAGCTTCTGGGCTGTGCCATGCGTACATTCC 105113
|||||
QY 1644 ttattctctcgtgtttccagaaacttaaggagaggaacacatttttggtgagtgattataacc 1703
|||||
Db 105114 TTATTCTTCTCGTGTTCAGAACCTTAAGGAGAGGCACATTTTGGTTGAGTGATTATAACC 105173
|||||
QY 1704 ctaggagccatgggtagctgctgattgcaggaacaaactcctcaacttctcgtccctgatgga 1763
|||||
Db 105174 CTAGGGACCATGGGTAGTGTGATGTCAGGAACAACTCCTCAACTTCTCTGGCCCTGATGGA 105233
|||||
QY 1764 ttaaggagagaggtacttacagatttatttcttcgtgtggactactgtccccagcatgaata 1823

Db 105234 TTAAGGAGAGGTTACTTACAGGTTATTCTTCGCTGTGGACTACTGTCCCAGCATGAATA 105293
|||||
QY 1824 gggcatcattattgaattattttgacaggaaggagactgggtgtatgctgcacagtaataa 1883
|||||
Db 105294 GGGCATCATTTATTAATTTATTTTCACAGGAAGGAGACTGGTGTATGCTGCACAGTAATAA 105353
|||||
QY 1884 tgtattacatgtctcagagttttaccaagcacotctgtgtgttttttgcotttggttat 1943
|||||
Db 105354 TGTATTTACATGTGTACAGAGTTTACCAAGCACCTCTGTGTGTGTGTGTGTGTGTGTAT 105413
|||||
QY 1944 tacactt-----gggacaaatttttaaaatttatatacatgcagagact 1985
|||||
Db 105414 TACACTTGGCCCTTGGGGGAGGTGGGCAAAATTTTAAATAATTTATACATGCAGAGACT 105473
|||||
QY 1986 gcagcgagagaagctaaagagacttgccttgcacacacagccagtggtgagagcctgaac 2045
|||||
Db 105474 GCAGGCGAGAGAAGCTAAGAGACTTGGCCCTGCCACACAGCCAGTGGTAGAGCCTGAAC 105533
|||||
QY 2046 tcaaacccaggtctcatctcaactcaggggctgttttcccaatcgctgtattgttcttaa 2105
|||||
Db 105534 TCAAAACCAGGCTCATCTCACCTCAGGGGCTGCTTTCCCCATCGCTGTATTGTCTTAA 105593
|||||
QY 2106 agtgatgggtgactagcaatgaagttaattctctagaaagcatgacaaatttcccttc 2165
|||||
Db 105594 AGTGATGGGTGACTAGGCAATGAAGTAATTTCTTAGGAAAGCATGACCAATTTCCCTTTC 105653
|||||
QY 2166 tccacctccctcttttctctccacccctcccccactcagcccccataatatatgcccacaa 2225
|||||
Db 105654 TCCACCTCCCTCTTTTCTCCACCCCTCCCCCATCAGCCCATATATATATGCCCCAATC 105713
|||||
QY 2226 tccaaagccttgctgctgctgcaaacctttactcttgaatgacttccacgctggagac 2285
|||||
Db 105714 TCCAAAGGCGTGTCTGTCCCTGCAAAAGCTTTACTTCTGAAATGACTTCCACGGCTGGAC 105773
|||||
QY 2286 g 2286
Db 105774 G 105774

RESULT 3
AC023534/c
LOCUS
DEFINITION Homo sapiens clone RP11-462N18, WORKING DRAFT SEQUENCE, 41
AC023534 229302 bp DNA linear HTG 04-MAY-2000
Homo sapiens clone RP11-462N18, WORKING DRAFT SEQUENCE, 41
unordered pieces.
AC023534
VERSION AC023534.3 GI:7705140
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 229302)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-462N18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 229302)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Chapel,V., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenestral,J., Ferreira,P., Fitzhugh,W., Forrester,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A.,
 Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
 Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
 Zody, M.

TITLE JOURNAL

Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 4, 2000 this sequence version replaced gi:7143452.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5244

Center clone name: 462.N.18

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 194690 bases at least Q40

Consensus quality: 207622 bases at least Q30

Consensus quality: 217061 bases at least Q20

Insert size: 20000; agarose-fp

Insert size: 223302; sum-of-contigs

Quality coverage: 4.5 in Q20 bases; agarose-fp

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1108: contig of 1108 bp in length
 * 1109 1208: gap of 100 bp
 * 1209 2504: contig of 1296 bp in length
 * 2505 2604: gap of 100 bp
 * 2605 3688: contig of 1084 bp in length
 * 3689 3788: gap of 100 bp
 * 3789 4962: contig of 1174 bp in length
 * 4963 5062: gap of 100 bp
 * 5063 6375: contig of 1313 bp in length
 * 6376 6475: gap of 100 bp
 * 6476 7673: contig of 1198 bp in length
 * 7674 7773: gap of 100 bp
 * 7774 9425: contig of 1652 bp in length
 * 9426 9525: gap of 100 bp
 * 9526 10997: contig of 1472 bp in length
 * 10998 11097: gap of 100 bp
 * 11098 12740: contig of 1643 bp in length
 * 12741 12840: gap of 100 bp
 * 12841 14467: contig of 1627 bp in length
 * 14468 14567: gap of 100 bp
 * 14568 16980: contig of 2413 bp in length
 * 16981 17080: gap of 100 bp
 * 17081 19176: contig of 2096 bp in length
 * 19177 19276: gap of 100 bp
 * 19277 21982: contig of 2606 bp in length
 * 21983 21982: gap of 100 bp
 * 21983 23826: contig of 1844 bp in length
 * 23827 23926: gap of 100 bp
 * 23927 26561: contig of 2635 bp in length
 * 26562 26661: gap of 100 bp
 * 26662 28846: contig of 2185 bp in length
 * 28847 28946: gap of 100 bp
 * 28947 32495: contig of 3549 bp in length

* 32496 32595: gap of 100 bp
 * 32596 34950: contig of 2355 bp in length
 * 34951 35050: gap of 100 bp
 * 35051 38193: contig of 3143 bp in length
 * 38194 38293: gap of 100 bp
 * 38294 41862: contig of 3569 bp in length
 * 41863 41962: gap of 100 bp
 * 41963 45898: contig of 3936 bp in length
 * 45899 45998: gap of 100 bp
 * 45999 50904: contig of 4906 bp in length
 * 50905 51004: gap of 100 bp
 * 51005 55346: contig of 4342 bp in length
 * 55347 55446: gap of 100 bp
 * 55447 61336: contig of 5890 bp in length
 * 61337 61436: gap of 100 bp
 * 61437 65806: contig of 4370 bp in length
 * 65807 65906: gap of 100 bp
 * 65907 71648: contig of 5742 bp in length
 * 71649 71748: gap of 100 bp
 * 71749 74886: contig of 3138 bp in length
 * 74887 74986: gap of 100 bp
 * 74987 79878: contig of 4892 bp in length
 * 79879 79978: gap of 100 bp
 * 79979 84280: contig of 4302 bp in length
 * 84281 84380: gap of 100 bp
 * 84381 89673: contig of 5493 bp in length
 * 89674 89773: gap of 100 bp
 * 89774 97168: contig of 7195 bp in length
 * 97169 97268: gap of 100 bp
 * 97269 101906: contig of 4638 bp in length
 * 101907 102006: gap of 100 bp
 * 102007 109439: contig of 7433 bp in length
 * 109440 109539: gap of 100 bp
 * 109540 117592: contig of 8053 bp in length
 * 117593 117692: gap of 100 bp
 * 117693 126037: contig of 8345 bp in length
 * 126038 126137: gap of 100 bp
 * 126138 135140: contig of 9003 bp in length
 * 135141 135240: gap of 100 bp
 * 135241 146384: contig of 11144 bp in length
 * 146385 146484: gap of 100 bp
 * 146485 160352: contig of 13868 bp in length
 * 160353 160452: gap of 100 bp
 * 160453 173458: contig of 13006 bp in length
 * 173459 173558: gap of 100 bp
 * 173559 191605: contig of 18047 bp in length
 * 191606 191705: gap of 100 bp
 * 191706 229302: contig of 37597 bp in length.

FEATURES

source

1..229302
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-462N18"
 /clone_lib="RPC1-11 Human Male BAC"

misc_feature

1..1108
 /note="assembly_fragment"

1209..2504
 /note="assembly_fragment"

2605..3688
 /note="assembly_fragment"

3789..4962
 /note="assembly_fragment"

5063..6375
 /note="assembly_fragment"

6476..7673
 /note="assembly_fragment"

7774..9425
 /note="assembly_fragment"

9526..10997
 /note="assembly_fragment"

11098..12740
 /note="assembly_fragment"

12841..14467
 /note="assembly_fragment"

```
/note="assembly_fragment"
14568..16980
/note="assembly_fragment"
17081..19176
/note="assembly_fragment"
19277..21882
/note="assembly_fragment"
21983..23826
/note="assembly_fragment"
23927..26561
/note="assembly_fragment"
26662..28846
/note="assembly_fragment"
28947..32495
/note="assembly_fragment"
32596..34950
/note="assembly_fragment"
35051..38193
/note="assembly_fragment"
38294..41862
/note="assembly_fragment"
41963..45898
/note="assembly_fragment"

Query Match      96.6%; Score 2208.6; DB 2; Length 229302;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 2253; Conservative 0; Mismatches 9; Indels 19; Gaps 2;

QY 25 gactcgatcaccttttgaacccaggctgcctgcctcccaagcttactactataactaga 84
Db 158375 GGCTGATACCTTTTGAACCCAGGCTGCCTGCCCTCCAAAGCTTGACTCATAACTAGA 158316

QY 85 ttctcaactgatgttgggccaagttctcaggttctctctctcagaccttctcttgaa 144
Db 158315 TTCTCAACTGATGTTGGGCCAAGGTTCTTAGGTTCTCTCTTGACCTTCTCTGAAGTA 158256

QY 145 ataagtctgataagctcatcgagagctgagccaggaacaatgttttcctgaactatc 204
Db 158255 ATAACTGATATAGTATCCCTCAGACAGAGTGAGCTACTCAGATCCCAAGGTTACCTTG 158196

QY 205 catgttatatgattctctcctcagacagagtgagctactcaagatcccaaggtgtaccctg 264
Db 158195 CATGTTATATGATTCTCTCAGACAGAGTGAGCTACTCAGATCCCAAGGTTACCTTG 158136

QY 265 agccagccaagtgatcatgacctatgctctctgttccagcctgccctttaacagct 324
Db 158135 AGCCAGCCAAGGTGATCCATGACCTCATGCTCTGTCTTCCAGCCTGCCCTTTAAACGCT 158076

QY 325 catccacacctgctgcccctcccgccttatctgcagacagtagtctaggatttcaagctacc 384
Db 158075 CATCCACCTGCTGCTGCCCTCCCGCCTATCTCGACACAGTAGTCTTAGGATTTACGCTGCC 158016

QY 385 ctgggggctcaatttccctctcagcttctgttagctgtctcgtcctgcctccactacc 444
Db 158015 CTGGGGGCTCATTTTCCCTCTCAGCTTCCGTTAGCTGTCTCCTGCTCCTCCACTCACC 157956

QY 445 tatctctcagcactctcaactcgtctcttcttctgtctcactcaatgctctcttgacatct 504
Db 157955 TATTAATCCAGCACTCTCACCCTGGCTCTCTTTCTCTCTCATFCACTGGCTCTTTGACATCT 157896

QY 505 ttatctcatagtagttagtggttcttggtaatgcctcaataatccacatggtgggaag 564
Db 157895 TTATCTCATAGTAGTAGTTAGGGGTTCTTGCTAATGCGCCTAATATCCACATGGTGGGAAG 157836

QY 565 gggggagtgggggaagagagtcgctgtgggctgtgctactcttctgaggggtaagactc 624
Db 157835 GGGGGAGTGGGGGAAGAGAGTGCCTGTGGGCTGTGCTACTTCTGAGGGTAAGACTC 157776

QY 625 gggccctccaggaacaaggattcagagctgtggcagctatagcacaagcagactgctggc 684
Db 157775 GGGCCCTCCAGGAACAAGGATTACAGGTTACAGGCTGGTGGCAGCTATAGCCAAAGCACTGCTGGC 157716

QY 685 cagggtatgcaaaaggagtattttgtgttaagaaaaataaacacactgagtagatgagat 744
```

```
||||| CAGGGATTGCAAGGAGTATTTTGTGCTTAAGAAAAATAAACAACTAGTATGAGAT 157715
QY 745 ggaaggaggggggtgttgggtgcagagagattgggaagagctctgccaaagggtgtgtcttac 804
||||| GAGGGAGGGGGGTGTTGGTCCAGAGAGAGATTGGGAAGAGTCTGCCAANGGTGTGTTCTAC 157655
QY 805 tcaactctctcttcttcttctcactgcactgcactgcactgcactgcactgcactgcactgcact 864
Db 157595 TCACTCTCTCTTTTCTTTCTATCTCCACTGAGCTGAGGAGCAGTTATCCTGTGCCACGT 157536
QY 865 cacattcctactcccgcttcccatgcctgacccaggctgggcaaacactcttctctgtaag 924
Db 157535 CACATTTCCTACTCTCCCTTTCCCATGCTCGACCCAGGTTGGGCAAACTCTTCTGTGTAAG 157476
QY 925 aaccagacaggaactatttttaggctctgtgccaataggtctcagtcacaactactcat 984
Db 157475 AACCCAGACAGGAACATATTTTAGGCTCTGTGTGCCATATGCTCAGTCAACAATCTCAT 157416
QY 985 ctctgcctctgtagcagcaaaagcaattagcaacaatatgtccaaacaatatgtgacccc 1044
Db 157415 CTCTGCCTCTGTAGCACGAAAGCAATTAGCAACAATATGTCAACAACAATATGTGACCCC 157356
QY 1045 atgaacaactttatttattatgatacggaacactctgaaataaatgtctctcttcttctt 1104
Db 157355 ATGAAAACCTTTATTTATTTATGATAGGAAACCTCGAAAATATGCTCTTTCTTTGATTTT 157296
QY 1105 ttccccaatcatataaaaaacgtaaaaaactactcttagtgcgaaggttaagcattctca 1164
Db 157295 TTCCCCCAATCATTTAAAAACCTTAAAAACTACTCTTTAGGTGCGCAAGGTTAAGCATCTCA 157236
QY 1165 gcttagcagtgccagcgtggatttgctgtgacctacagttggtggccaatctccctgattccc 1224
Db 157235 GCTTAGCAGTGGCAGGCTGGATTGGCTTGTGACCTACAGTTGGCCAACTCCCTGATTCCC 157176
QY 1225 aaaaagtattctcctcaggagtggtgggcaaaactacttatgggaagtgctggtataaacagagt 1284
Db 157175 AAAATGTTATTCCTCAGGGATGTGGGCAAACTACTATGGGAAATGCTGATTAACAGAGT 157116
QY 1285 taagaagcatcagacatttccagagccgggctagacacatgcagggctctctcaactgaacct 1344
Db 157115 TAAGAAGCATCAGACATTTCCAGGACGGGCTAGACATGCCAGGGCTCTCTTAACCTGAGCT 157056
QY 1345 catt-ggattcatctgttctcaggaggtcttgcgaagaagaaattcctcaaacctagag 1403
Db 157055 CATTAGGATTTCATCTGTTCATGGAGGATCTTGCAAGACAAAGAAATTCCTCAAACTAGAG 156996
QY 1404 tctgaggactgtcttgggaaacactgctctgtgtgagcctcactgcctcactgcggcacatgta 1463
Db 156995 TCTGAGGACTGTGCTTTTGGGAAACACTGCTCTGCTTGTATGGCCCTCACTGGGCACATGGTA 156936
QY 1464 gaaactagagctgagtgcttgccttagctggagatagggctcagagctcttgcactgacctgg 1523
Db 156935 GAATCTAGAGCTGAGTGCTTGTGCTAGCTGGAGATAGGGCTCAGAGCTCTTGACTGCCCTGG 156876
QY 1524 cagtcttgacacatacaogctgtctgttccctgagttggttcagagccacacagggccaag 1583
Db 156875 CAGCTTTGACACATCAGCTGCTGTGTGCTCCCTCAGTGGTTTCAGAGCCACACAGGCCAAG 156816
QY 1584 actagccacacagagaccaggcctccagcttctctgggcttgcctatgcagcaatttcc 1643
Db 156815 ACTAGCCACCACAGAGCACACAGGCTTCCAGCTTCTGGGCTTGTCCATGGCTATGATTTCC 156756
QY 1644 ttattcttctctgggtttccagaacctaaaggagagcacatttttggttagtgattataacc 1703
Db 156755 TTATTCTCTCTGGTTTCCAGAACCTTAGGAGAGCACATTTTGGTTGAGTGATTATTAACC 156696
QY 1704 ctagggaacctggttagctgcatgtcaggaaacactcctcaactctcctggtccctgatgga 1763
Db 156695 CTAGGGACCATGGGTAGCTGCATGTCAAGAAACACTCCTCAACTTCTCGCCCTGATGGA 156636
QY 1764 ttaaagagagaggtacttacaagggtatttcttctgctgtggactactgtccccagatgata 1823
|||||
```


Qy	1984	ctgacg	cgacgagaaagctaagagacttgcacctgcacacagccagtggtagagcctga	2043
Db	82683	CTGCACGCGGAAGAAGACTTGAAGAGACTTGCCTGCGCCACAGCAGCAGTGGTAGAGCCCTGA	82742	
Qy	2044	acccaacccaggtctcatctcaectcaagggtgcttccccatcgctgtattgtcctt	2103	
Db	82743	ACTCAAAACCCAGGTCTCATCTCACTCAGGCGCTGCTTCCCATCGCTGTATTGTCTCTT	82802	
Qy	2104	aaagtcatggtgactagggcaatgaagtaattctctagaaagcatgacccaatttcctct	2163	
Db	82803	AAAGTGATGGGTGACTAGGCAATGAAGTAATTTCTTAGGAAGCATGACCAATTTCCCTT	82862	
Qy	2164	tctccacctccctctttttctccaccctccctcccatcagcccatatataatccccaaa	2223	
Db	82863	TCFCCACCTCCCTCTTTTTCCTCCACCCCTCCCATCAGCCCATATATATGCCCAAA	82922	
Qy	2224	tctcccaaaagctgctgctgcaaacctttactctgaaatgaactccaaggtggg	2283	
Db	82923	TCFCCAAAGCCTTGCTTGCTGCAAAACCTTTACTTCTGAAATGACTTCCACGGCTGGG	82982	
Qy	2284	acg	2286	
Db	82983	ACG	82985	
RESULT 5				
AY062931 11114 bp DNA linear PRI 26-NOV-2001				
LOCUS Homo sapiens interleukin 24 (IL24) gene, complete cds.				
DEFINITION AY062931				
ACCESSION AY062931				
VERSION AY062931.1 GI:17104816				
KEYWORDS human.				
SOURCE Homo sapiens				
ORGANISM Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE 1 (bases 1 to 11114)				
AUTHORS Rieder,M.J., Armel,T.Z., Carrington,D.P., Chung,M.-W., Lee,K.L.,				
Poel,C.L., Toth,E.J., Yi,Q. and Nickerson,D.A.				
TITLE Direct Submission				
JOURNAL Submitted (13-NOV-2001) Molecular Biotechnology, University of				
Washington, 1705 NE Pacific, Seattle, WA 98195, USA				
COMMENT To cite this work please use: SeattleSNPs: NHLBI Program for				
Genomic Applications, UW-FHCRC, Seattle, WA (URL:				
http://pga.mbt.washington.edu).				
FEATURES				
source				
1..11114				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
repeat_region				
543..856				
/rpt_family="MERL_type"				
/rpt_type="dispersed				
variation				
627..634				
/frequency="0.01"				
/replace="a"				
variation				
634				
/frequency="0.10"				
/replace="c"				
variation				
806				
/frequency="0.05"				
/replace="g"				
repeat_region				
884..925				
/rpt_family="MERL_type"				
/rpt_type="dispersed				
variation				
1619..1620				
/frequency="0.07"				
/replace="n"				
repeat_region				
1625..1759				
/rpt_family="SINE"				
/rpt_type="dispersed				
variation				
1627				
/frequency="0.02"				
/replace="c"				

Qy 1263 gaagtgctggattaaacagagattaaagacatcagacattccaggacggctagcacat 1322
Db 901 CAATGCTGGATTAAACAGAGTTAAGAGCATCAGACATTTCCAGGACGGCTAGCACAT 960
Qy 1323 gccagggtctctcaactgaactcaatt-ggattcaatctgtttcagagatcttgcaaga 1381
Db 961 GCCAGGGCTCTTAAGTGAAGCTATTAAGGATTCATCTGTTTCATGAGGATCTTCAAGA 1020
Qy 1382 caagaattcccaaaccttagagctgagagactgtgctttgggaacactgctctgtga 1441
Db 1021 CAAGAAATTCCTCAACCTAGAGCTGAGGACTGTGCTTTGGGAAACACATGCTGCTTCA 1080
Qy 1442 tgcctcaactgggacatgttagaatactagagctgagctgctgctgctgagatagagg 1501
Db 1081 TGCCCTCACTGGGACATATGAGTAATCTAGAGCTGAGTGTGCTTGTAGCTGGAGATAGG 1140
Qy 1502 tcaagactctgactgctgagctgtgagctgttgacacatacagctgtctgtgctcctgagtg 1561
Db 1141 TCAGAGCTCTTGAAGCTGGCAGTCTTGACACATCAGCTGTCTGTGCTCCTGAGTG 1200
Qy 1562 gttcagagccacacagggccaaagactagccacacagagacacagagctccagcttcttg 1621
Db 1201 GTTCAGAGCCACAGGCGCAAGACTAGCCACAGAGACACAGGCTCCAGCTTTCTGG 1260
Qy 1622 gctgtccatcgctacatttctcttctctgctgctgtttccagaaacctaaagagagcaca 1681
Db 1261 GCTTGTCCATGCGTACATTTCTTATCTTCTCTGCTGTTTCCAGAACCTTAAGGAGAGCACA 1320
Qy 1682 ttttggtgagtgattataacccctagggacacatgggtagctgctgctgctgctgctgctg 1741
Db 1321 TTTTGGTGTGAGTGATTAACCCCTAGGAGCATTGGGTAGCTGCTGATGCTGATGCTGCAAGAACACTCC 1380
Qy 1742 tcaactctctgctgctgagctgattaaagagagagcttactacaggttatttctcgtctg 1801
Db 1381 TCAACTTCTGCGCTGATGATTAAGAGAGAGTACTTACAGGTATTTCTTCGCTGTG 1440
Qy 1802 gactactgcccagcatgaatagggatcattatgataattatttgacaggaaggagact 1861
Db 1441 GACTACTGTCACGATGAATAGGCGATCATTAATTGAAATTAATTTTACAGGAAGAGAGACT 1500
Qy 1862 ggtgatctgcacagtaataatgtatttacatgtgtacagagcttaccagacacctcgt 1921
Db 1501 GGTGTATGCTGCACAGTAATAATGATTAATACATGTGTACAGATTTACCAGGACCTCTG 1560
Qy 1922 tttgttttttgccttctgtttattacatt-----gggacaaaattttt 1963
Db 1561 TGTGTTTGTGCTGCTGTTTATTACATTTGCCCTTGGGGGAGGTGGGACAAATTTTATA 1620
Qy 1964 aaaaattatacatgacagactgcagcgcagagagagctagagacttgccttgcctcaca 2023
Db 1621 AAAATTTATATGACAGAGACTGCAGCGCAGAGAGCTAAGAGACTTGCCTTGCCTCACA 1680
Qy 2024 cagccagtggttagagctgaaactcaaacccaggtctcatctcatcctcaggggctgtttc 2083
Db 1681 CAGCCAGTGTGTAGAGCTGAACCTCAAAACCCAGGTCTCATCTCACTCAGGGGCTGCTTC 1740
Qy 2084 cccatcgctgtattgtccttaagtgtgagtgactaggaatgaagtaattctctagga 2143
Db 1741 CCCATCGCTGTATTCTCTTAAAGTGATGGGTGACTAGGCAATGAAGTAATTTCTTAGGA 1800
Qy 2144 aagcatgaccaatttccctttctccactcctccttcttctccacccctcccccacag 2203
Db 1801 AAGCATGACCAATTTCCCTTTCTCCACTCTCTCTTTTCTCCACCTCTCCCTCCATCAG 1860
Qy 2204 ccccatatataatgcccaaatctccacaaagccttgctgtccttgcaaacctttactcttg 2263
Db 1861 CCCCCATATATGCCCAATCTCCACAAAGCCTTGTCTGCTGCTGCAAACTTTACTTCTG 1920
Qy 2264 aaatgacttccagcgtgggagc 2286
Db 1921 AAATGACTTCCACGGCTGGGAGC 1943

RESULT 6
AF276916
LOCUS Homo sapiens melanoma differentiation association protein 7 gene,
DEFINITION complete cds.
ACCESSION AF276916
VERSION AF276916.1 GI:11907920
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 7025)
AUTHORS Peat,J., Kube,D., Eskdale,J., Juelliger,S. and Gallagher,G.
TITLE The human MDA-7 gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 7025)
AUTHORS Peat,J., Kube,D., Eskdale,J., Juelliger,S. and Gallagher,G.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-2000) Department of Surgery, University of
Glasgow, Queen Elizabeth Building, Glasgow Royal Infirmary, Glasgow
G31 2ER, Scotland
FEATURES
Source
1..7025
/organism="Homo sapiens"
/db_xref="taxon.9606"
join(<1804..1847,3285..3480,4261..4323,5460..5618,
5964..6038,6942..>7025)
/product="melanoma differentiation association protein 7"
join(1804..1847,3285..3480,4261..4323,5460..5618,
5964..6038,6942..7025)
/codon_start=1
/product="melanoma differentiation association protein 7"
/protein_id="AAG41401.1"
/db_xref="GI:11907921"
translation="MNPQRLQSLWTLARPFCPPLLATASQMQMVLPCGLFTLLWS
QVGAQGEHFHGPCQKGVVPOKLEAFWAKDITMOADNITSARKLQEVQLNVSD
AESCYLVHTLLEFLYLVTKFNKHNRTVEVTLKSFSLANNFVLIVISQLOPQSEDMF
SIRDSAHRRFLFRRAFKQDLDEAALTALKEVDILLTWQRFYK1"
BASE COUNT 1759 a 1668 c 1759 g 1838 t 1 others
ORIGIN
Query Match 59.7%; Score 1365.2; DB 9; Length 7025;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1425; Conservative 0; Mismatches 14; Indels 22; Gaps 3;
Qy 848 taccctgtcccccacgtcacattcctactccgttcccatgctgctgacccaggtgggc 907
Db 1 TATCCTGTCCCCACGTACATTCCTACTCCGTTTCCCATGCTGACCCAGGTTGGGC 60
Qy 908 aaactctctctgtaagaaccagacaggaactattttaggctctgtgtgcaatatggtct 967
Db 61 AAACCTCTTCTCTGTAAGAACCAGACAGAACTATTTTAGCGCTGTGTGCATATGCTCT 120
Qy 968 cagtcacactactctctgctctgtagcagcaagaatagcaacaatatgtcaa 1027
Db 121 CAGTCACAACTACTACTACTCTGCTCTGTAGCAGCAAGCAATTAGCAACAATATGTCAA 180
Qy 1028 caaacatgtgacccatgaaactttattattatgatacggaaacctgaaataat 1087
Db 181 CAACATATGTGACCCCATGAAACTTTATTTATTTATGATACGAAACCTGAAATAAT 240
Qy 1088 gtcttcttttgattttttcccaatcattaaaaaacgtaaaaactactcttaggtcgca 1147
Db 241 GTCTTCTCTTTGATTTTTCCTCCCAATCATTAATAAAACGTAAAAACTACTCTTAGGTGCGCA 300
Qy 1148 aggttaagccattctcagcttagcagtgagcaggtggtgattggcttgacctacagttg 1207
Db 301 AGGTTAACCCATCTCAGCTTAGCAGTGGCAGGCTGGATTTGGCTTGTGACCTACAGTTG 360
Qy 1208 gccaatccctgattcccaaaatgtattcctcagggaatgtgggcaaacactatgtgggaagt 1267
|||||

Db 361 GCCAATCCCTGATTCCTCCAAATGTATTTCCTCAGGGATGTGGCAAAATACTTATGGGAAAT 420
QY 1268 gctgatttaaacagagtttaagagctacagacattccagagcgtctagacatgccag 1327
|||||
Db 421 GCTGATTAACACAGATTAGAGCATACAGCATTTCCAGGACGGCTAGACACATGCCAC 480
QY 1328 ggcctctaaatgcacctcatt-ggattcatctgtttcatctgaggagatcttgcgaagacaaga 1386
|||||
Db 481 GGCTCTCTAACTGACCTCAATTAGGATTCATCTCTTTCATGGAGGATCTTGCAAGACAAGA 540
QY 1387 attctcaaacctagaatcgaagactgctcttggaaacactgctctgctgatgcc 1446
|||||
Db 541 ATTCTCTAAACCTAGAGCTGAGACATGCTTTGGGAACACTGCTCTGCTGTATGCC 600
QY 1447 tcaactgggacacatggtagaatctagagctgagtgccctgtctgctgagtgagagtagggtcaga 1506
|||||
Db 601 TCACCTGGGCACATGTTAGATCTAGAGCTGAGTGCCTTCTAGCTGGAGATAGGCTCAGA 660
QY 1507 gctcttgactcctgacgtcttggacacatacagctgctctgtctgctcctcctgagtggtcca 1566
|||||
Db 661 GCTCTTGACTCCCTGGCAGCTTTGACATACATCAGCTGTCTGTGCTCCCTGAGTGGTTCA 720
QY 1567 gagcacacagccaaagactagccacacagagcaccagcgcctccagctcttctgggcttg 1626
|||||
Db 721 GAGGCACACAGCCCAAGACTAGCCACACAGACACAGCCCTCCACAGCTTCTGGGCTTG 780
QY 1627 tccatgcgtacattctcttctctctggtttccagaaacctaaagagagacacatttg 1686
|||||
Db 781 TCCATGCGTAGATTTCTCTTCTGCTTTCCAGAACCTTAGAGAGAGGACACATTTTG 840
QY 1687 gttgagtgattataacctaggaccatgggtagctgctgctcaggaacactcctcaac 1746
Db 841 GTTGAGTGATTAACCTTAGGACCATGGGTAGCTGCTATGTGCTAGGTAATTTTACGGA 900
QY 1747 tctctggcctcgtggttaaaagagaggtacttacagattattcttcgctgt---gga 1803
Db 901 TTCTGGCCCTGAGGATTAAGAGAGAGGTACTTACAGGTTATTTCTGCTATTACGGA 960
QY 1804 ctactgtccacagatgaatgggacatcaattatgaattatttgacgaagagagactgg 1863
Db 961 CTACTGTCCACAGCATGAATAGGCGCATCATTTATGAATTTATTTGACGAAGGAGACTGG 1020
QY 1864 tgaatgcgcacagataaatagtattacatgtgtacagatttaccagacacctctgt 1923
Db 1021 TGTATGCTGCACAGATATATGTTATTTACATGTGTACAGAGTTTACCAGACACCTCTGTG 1080
QY 1924 tggtttggcttgyttttatcacctt-----gggacaaaatttttaa 1965
Db 1081 TTGTTTTTGGCTCTGTTTATTACACTTGCCCTTGGGGGAGGTGGGACAAAATTTTTTAA 1140
QY 1966 aattatcacatgcagagactgcgcagagagctaaagagacttgccttgcctccacaca 2025
Db 1141 AATTTATACATGCAGAGACTGCAGCGCAGAGAGCTTAGAGACTTGGCCCTGCCACACACA 1200
QY 2026 gccagtgtagagctgaactcaaacccaggtctcatctcaactcaggggctgctttccc 2085
Db 1201 GCCAGTGTAGAGCTCACTCAAAACCCAGGCGCTNATCTCACTCAGGGGCTGCTTTCCC 1260
QY 2086 catcgtgattgtccttaagtagtggtgactaggaatgaatgaattctctaggaaa 2145
Db 1261 CATCGCTGAATTTGCCCTTAAGTGATGGGTGACTAGGCAATGAAGTAAATCTCTAGGAAA 1320
QY 2146 gcataaccaatttcccttctccactcctctctcttctccactcctcccccacacagcc 2205
Db 1321 GCATGACCAATTTCCCTTCTCCACCTCCCTCTTTTCTCCACCCCTCCCCCATCAGCC 1380
QY 2206 cccatatatgcccaatccacaaagccttgcctgctgcacaaaccttacttctgaa 2265
Db 1381 CCCATATATATGCCCAATCTCCACAAAGCCTTGCTTGCTGCTGCAAAACCTTTACTTCTGAA 1440
QY 2266 atgacttccacgctggagacg 2286
|||||
Db 1441 ATGACTTCCACGGCTGGGACG 1461

RESULT 7

HSR243947
LOCUS HSR243947 29034 bp DNA linear PRI 09-NOV-2000
DEFINITION Homo sapiens VCX-A gene for variably charged X chromosome protein.
ACCESSION AJ243947
VERSION AJ243947.1 GI:5912555
KEYWORDS VCX-A gene.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 29034)
AUTHORS Li,X.M., Yen,P.H. and Shapiro,L.J.
TITLE Characterization of a low copy repetitive element S232 involved in the generation of frequent deletions of the distal short arm of the human X chromosome
JOURNAL Nucleic Acids Res. 20 (5), 1117-1122 (1992)
MEDLINE 92195814
REFERENCE 2 (bases 1 to 29034)
AUTHORS Fukami,M., Kirsch,S., Schiller,S., Richter,A., Benes,V., Franco,B., Muroya,K., Rao,E., Merker,S., Niesler,B., Ballabio,A., Ansorge,W., Ogata,T. and Rappold,G.A.
TITLE A member of a gene family on xp22.3, VCX-A, is deleted in patients with x-linked nonspecific mental retardation
JOURNAL Am. J. Hum. Genet. 67 (3), 563-573 (2000)
MEDLINE 20395828
REFERENCE 3 (bases 1 to 29034)
AUTHORS Richter,A.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1999) Richter A., Biochemical Instrumentation, EMBL, Meyerhofstrasse 1, Heidelberg, 69117, GERMANY
FEATURES
source 1. 29034
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="between 40BT and DXS139"
/clone="LLNC110"
/sex="male"
/cell_line="lymphoblastoid"
/clone_lib="Lawrence Livermore X-chromosome specific library"
/note="Variably charged, X chromosome mRNA on CRI-S232A"
10574..10732
/gene="VCX-A"
/number=1
10574..11947
/gene="VCX-A"
10733..11046
/gene="VCX-A"
/number=1
11047..11294
/gene="VCX-A"
/number=2
join(11195..11294,11487..11947)
/gene="VCX-A"
/codon_start=1
/product="VCX-A protein"
/protein_id="CAB56197.1"
/db_xref="GI:5912556"
/translaton="MSPKPRASGPPAKATEAGKRKSSQSPSPDPKKTKTKVAKGKA VRGRGKGAATKMAATAPAEAGSPPAAGPSQSPQLPOHLPPEPVSQTHD PLQSELEPLSQSEVEEPLSQSEVEEPLSQSEVEEPLSQSEVEEPLSQSEV EEPLOSQVEEPLSQSEMEELPSV"
11295..11486
/gene="VCX-A"
/number=2
11487..11947
/gene="VCX-A"
/number=3
11697..11936
repeat_region

Thu Jun 6 16:21:04 2002

repeat_region	12097..12297	/citation=[1] /rpt_family="RUI" /citation=[1] /rpt_family="RUI"	9123 a 5623 c 5586 g 8631 t 71 others	Query Match Best Local Similarity 4.0%; Score 91.8; DB 9; Length 29034. Matches 114; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
BASE COUNT	9123 a 5623 c 5586 g 8631 t 71 others			
ORIGIN				
LOCUS	AC108684/c			
DEFINITION	Human sapiens chromosome Xp clone RP11-359020, WORKING DRAFT			
VERSION	AC108684.1			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE	human.			
ORGANISM	Human sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.J., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbara, J., Benton, J., Binage, K., Blankenbrow, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, R.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgeson, A., Hognes, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, Y., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuogu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Roife, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Sizson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, K., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,			
LOCUS	AC108684			
DEFINITION	Human sapiens chromosome Xp clone RP11-359020, WORKING DRAFT			
VERSION	AC108684.1			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE	human.			
ORGANISM	Human sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.J., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbara, J., Benton, J., Binage, K., Blankenbrow, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, R.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgeson, A., Hognes, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, Y., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuogu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Roife, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Sizson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, K., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,			
LOCUS	AC108684			
DEFINITION	Human sapiens chromosome Xp clone RP11-359020, WORKING DRAFT			
VERSION	AC108684.1			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE	human.			
ORGANISM	Human sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.J., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbara, J., Benton, J., Binage, K., Blankenbrow, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, R.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgeson, A., Hognes, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, Y., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuogu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Roife, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Sizson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, K., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,			
LOCUS	AC108684			

Qy 1011 tagacaacaatgtgcaacaaacatatgtgac 1041
 | ||||| || || |||||
 Db 159298 CACAGACAATGTGTAATAATTAATAATGTGC 159268

RESULT 9
 AC002090/c 137769 bp DNA linear PRI 06-JUN-1998
 DEFINITION Genomic sequence from Human 17, complete sequence.
 ACCESSION AC002090
 VERSION AC002090.1 GI:2160130
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 137769)
 AUTHORS Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C. and Lander,E.S.
 TITLE Genomic sequence from Human 17
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 137769)
 AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
 Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
 Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Frapp,W.J.,
 Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
 Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
 Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
 Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
 Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (13-MAY-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 3 (bases 1 to 137769)
 AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
 Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
 Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Frapp,W.J.,
 Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
 Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
 Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
 Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
 Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 4 (bases 1 to 137769)
 AUTHORS Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C., Lander,E.S.,
 McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
 Daly,M.J., Devon,K., Dewar,K., Forreest,C., Gage,D., Geraigery,K.,
 Hagos,B., Harris,K., Huang,J., Hui,L., Jacotot,L., Kirby,A.,
 Lane,M., Mackenzie,J., Marquis,N., McDermott,J., Molla,M.,
 Morrow,J., Nachman,A., Naylor,J., Nussbaum,C., O'Connor,T.,
 Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G.,
 Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
 Wilmer,F., Zemtseva,I. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 5 (bases 1 to 137769)
 AUTHORS Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C., Lander,E.S.,
 McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
 Daly,M.J., Devon,K., Dewar,K., Forreest,C., Gage,D., Geraigery,K.,
 Hagos,B., Harris,K., Huang,J., Hui,L., Jacotot,L., Kirby,A.,
 Lane,M., Mackenzie,J., Marquis,N., McDermott,J., Molla,M.,
 Morrow,J., Nachman,A., Naylor,J., Nussbaum,C., O'Connor,T.,
 Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G.,
 Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
 Wilmer,F., Zemtseva,I. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (09-SEP-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 6 (bases 1 to 137769)
 AUTHORS Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C., Lander,E.S.,
 McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,

Daly,M.J., Devon,K., Dewar,K., Forreest,C., Gage,D., Geraigery,K.,
 Hagos,B., Harris,K., Huang,J., Hui,L., Jacotot,L., Kirby,A.,
 Lane,M., Mackenzie,J., Marquis,N., McDermott,J., Molla,M.,
 Morrow,J., Nachman,A., Naylor,J., Nussbaum,C., O'Connor,T.,
 Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G.,
 Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
 Wilmer,F., Zemtseva,I. and Zody,M.

TITLE
 JOURNAL
 COMMENT

Submitted (06-JUN-1998) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 4, 1997 this sequence version replaced gi:2085774.
 The Staden databases, finishing information, and all
 chromatographic files used in the assembly of this clone are
 available from our anonymous ftp site.

All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.

FEATURES
 source

1. 137769
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /clone="4314"
 /clone_lib="Research Genetics/Cal Tech CITB9785K-B (plates
 1-194)"
 repeat_region 9. 113
 /rpt_family="MLTIC"
 repeat_region 114. 416
 /rpt_family="AluX"
 repeat_region 418. 860
 /rpt_family="MLTIC"
 repeat_region 951. 1021
 /rpt_family="(CATA)n"
 repeat_region complement(1659. 1701)
 /rpt_family="(CA)n"
 repeat_region 2009. 2282
 /rpt_family="LINE2"
 repeat_region complement(2599. 2700)
 /rpt_family="MIR"
 repeat_region complement(2709. 2884)
 /rpt_family="LINE2"
 repeat_region complement(3877. 4032)
 /rpt_family="LINE2"
 repeat_region 4630. 4841
 /rpt_family="MIR"
 repeat_region complement(5998. 6079)
 /rpt_family="(GA)n"
 repeat_region complement(7528. 7829)
 /rpt_family="AluSX"
 repeat_region complement(8692. 8879)
 /rpt_family="MIR"
 repeat_region 9321. 9667
 /rpt_family="LINE2"
 repeat_region complement(9688. 9982)
 /rpt_family="AluSg"
 repeat_region 9985. 10210
 /rpt_family="LINE2"
 repeat_region 11052. 11216
 /rpt_family="MIR"
 repeat_region complement(12296. 12393)
 /rpt_family="(GGGA)n"
 repeat_region 12572. 12644
 /rpt_family="AT-rich"
 repeat_region 12908. 12995
 /rpt_family="LINE2"
 repeat_region complement(14208. 14363)
 /rpt_family="LINE2"
 repeat_region complement(14577. 14686)
 /rpt_family="LINE2"
 repeat_region 14837. 14904
 /rpt_family="(CAGC)n"
 repeat_region complement(14906. 14963)

repeat_region /rpt_family="LINE2"
repeat_region complement(17782..17841)
repeat_region /rpt_family="AT_rich"
repeat_region complement(19230..19258)
repeat_region /rpt_family="AT_rich"
repeat_region complement(19446..19742)
repeat_region /rpt_family="AluY"
repeat_region 19749..19785
repeat_region /rpt_family="AT_rich"
repeat_region complement(20413..20518)
repeat_region /rpt_family="MER5B"
repeat_region 21555..21649
repeat_region /rpt_family="MER5A"
repeat_region complement(21855..21892)
repeat_region /rpt_family="LINE2"
repeat_region complement(22871..22908)
repeat_region /rpt_family="AT_rich"
repeat_region complement(23129..23168)
repeat_region /rpt_family="LINE2"
repeat_region complement(23296..23654)
repeat_region /rpt_family="LINE2"
repeat_region 23894..24398
repeat_region /rpt_family="MLT2CA"
repeat_region complement(24536..24675)
repeat_region /rpt_family="MIR"
repeat_region complement(24946..25005)
repeat_region /rpt_family="MIR"
repeat_region 26012..26140
repeat_region /rpt_family="LINE2"
repeat_region 26342..26477
repeat_region /rpt_family="MIR"
repeat_region 26712..26929
repeat_region /rpt_family="MER58A"
repeat_region complement(27265..27327)
repeat_region /rpt_family="MIR"
unsure complement(28097..28886)
/note="Sequence in this region could not be obtained from the other strand or with an alternative sequencing chemistry."
repeat_region 28421..28587
repeat_region /rpt_family="MIR"
repeat_region 29156..29359
repeat_region /rpt_family="L1"
repeat_region 30191..30324
repeat_region /rpt_family="MIR"
repeat_region 30411..30728
repeat_region /rpt_family="MER7A"
repeat_region complement(30998..31061)
repeat_region /rpt_family="LINE2"
repeat_region complement(31629..31659)
repeat_region /rpt_family="AT_rich"
repeat_region 33484..34739
repeat_region /rpt_family="L1R12"
repeat_region 34740..35075
repeat_region /rpt_family="THE1B"
repeat_region complement(35105..35297)
repeat_region /rpt_family="MIR"
repeat_region complement(35585..35766)
repeat_region /rpt_family="MIR"
repeat_region complement(36289..36589)
Query Match 4.0% Score 91.2; DB 9; Length 137769;
Best Local Similarity 68.5%; Pred. No. 2.7e-14;
Matches 126; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 912 ttctctgtaagaccgacaggaactatttttaggtctctgtgccaataggtctcagt 971
Db 44990 TTTTCCAGTGAAGGCCAGATAGTAACATGTTAGGTCTCGTGGGCCCATATGGTCTCAGT 44931
QY 972 cacaactactatctctcctctctagcagcaaatagcaacaaatagcaacaaa 1031
Db 44930 TGCNACTACTCAACTGCCCTTGTGTGGGNAAGCAGTCACAGACATATATGCAACAAA 44871

QY 1032 catatgtaccccatgaaactttattattattgatacggaaacctgaaataatgtct 1091
Db 44870 TAGGGGTGACTGTTCCCAATGAACATTTGTACAGATGCTGAAAGTTGAAATTTATATAA 44811
QY 1092 ttct 1095
Db 44810 TTTT 44807
RESULT 10
AL139275 193126 bp DNA linear PRI 04-SEP-2001
LOCUS Human DNA sequence from clone RP11-552E20 on chromosome
DEFINITION 6p12.3-21.2, complete sequence.
ACCESSION AL139275
VERSION AL139275.30 GI:15485119
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 193126)
AUTHORS Smith,M.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Sep 6, 2001 this sequence version replaced gi:14800199.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em.: EMBL; SW.:
SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-552E20 is from the library RPC1-11.2 constructed by the group
of Piefer de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-552E20 The true
left end of clone RP11-552E20 is at 161600 in this sequence. The
true right end of clone RP11-200E19 is at 35574 in this sequence.
FEATURES
Location/Qualifiers
1..193126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p12.3-21.2"
/clone="RP11-552E20"
/clone_lib="RPC1-11.2"
12..455
/note="match: GSS: Em:AQ807014"
4253..4780
/note="match: GSS: Em:B71338"
4256..4678
/note="match: GSS: Em:AQ437688"
complement(8999..9566)
misc_feature
misc_feature
misc_feature
misc_feature

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 178477)
Biren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 14, clone RP11-354A24
Unpublished
2 (bases 1 to 178477)
Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bida, F.,
Boguslavsky, L., Boukigalier, B., Brown, A., Burkett, G., Castle, A.,
Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
Ferrelira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, J. C., Lebecky, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Olivari, T. M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 17, 2000 this sequence version replaced gi:6980319.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITB
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6051
Center clone name: 354_A24

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 172070 bases at least Q40
Consensus quality: 175121 bases at least Q30
Consensus quality: 176261 bases at least Q20
Insert size: 181000; agarose-fp
Insert size: 177277; sum-of-contings
Quality coverage: 5.4 in Q20 bases; agarose-fp
Quality coverage: 5.5 in Q20 bases; sum-of-contings

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1694 1793: gap of 1693 bp in length
* 1794 1793: gap of 100 bp
* 1794 4078: contig of 2285 bp in length
* 4079 4178: gap of 100 bp
* 4179 14243: contig of 10065 bp in length
* 14244 14343: gap of 100 bp
* 14344 21787: contig of 7444 bp in length
* 21788 21887: gap of 100 bp
* 21888 31116: contig of 9229 bp in length
* 31117 31216: gap of 100 bp
* 31217 39038: contig of 7822 bp in length
* 39039 39138: gap of 100 bp
* 39139 46809: contig of 6771 bp in length
* 46810 46909: gap of 100 bp
* 46910 57558: contig of 10649 bp in length
* 57559 57658: gap of 100 bp

* 57659 73890: contig of 16232 bp in length
* 73891 73990: gap of 100 bp
* 73991 96635: contig of 22645 bp in length
* 96636 96735: gap of 100 bp
* 96736 120417: contig of 23682 bp in length
* 120418 120517: gap of 100 bp
* 120518 149372: contig of 28855 bp in length
* 149373 149472: gap of 100 bp
* 149473 178477: contig of 29005 bp in length.

FEATURES

Source

1. 178477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14"
/clone="RP11-354A24"
/clone.lib="RP11-11 Human Male BAC"

misc_feature

1. 1693
/note="assembly_fragment"
1794..4078
/note="assembly_fragment"
4179..14243
/note="assembly_fragment"
14344..21787
/note="assembly_fragment"
clone_end:T7
vector_side:right

misc_feature

21888..31116
/note="assembly_fragment"
31217..39038
/note="assembly_fragment"
39139..46809
/note="assembly_fragment"

misc_feature

clone_end:SP6
vector_side:right
46910..57558
/note="assembly_fragment"

misc_feature

57659..73890
/note="assembly_fragment"
73991..96635
/note="assembly_fragment"

misc_feature

96736..120417
/note="assembly_fragment"
120518..149372
/note="assembly_fragment"
149473..178477
/note="assembly_fragment"

BASE COUNT 51444 a 37184 c 36608 g 52041 t 1200 others

ORIGIN

Query Match 3.9%; Score 90; DB 2; Length 178477;

Best Local Similarity 61.5%; Pred. No. 5.9e-14;

Matches 182; Conservative 0; Mismatches 105; Indels 9; Gaps 2;

Qy 861 acgtcacatctctactccgttcccatgctgagccaggttgcccaactcttctgt 920

Db 10108 AGGTTCTATTGCTAGCTTTGCTACAAATGACCAAGACAAAGGTCAGTAACTTTTCTCT 10049

Qy 921 aaagaaccagacaggaactatttttaggtctgtgcatatggtctcagtcacaaactac 980

Db 10048 GAAGGACTAGATAGTAAATATTTTCAGGCTTTGCAGACCATATGTCCTTTGCAACTAC 9989

Qy 981 tcattctgcctcttagcacgaagcaattagcaacaataatgtcaacaacatagtga 1040

Db 9988 TCAAGTCGCTGCTAGTATGAAGCCACCCTGGACAAATATGTAATGATGGGCATCA 9929

Qy 1041 ccccat--gaaacattttatttattggtatcggaacacctgaaataatgtcttctc--- 1094

Db 9928 CTGCATTCCAATAAATTTTCTTTATGACATTCAGATTTTGAATTTTACATGTCAGAA 9869

Qy 1095 ---ttttgattttttccccaatcattaaaaaacgtaaaaactactcttaggtcgca 1147

Db 9868 AATATTATTCGTTTCCCAATCATTAACATATAAAACCACTCTTAAGTCACA 9813

```

RESULT 13
AC098645
LOCUS
DEFINITION
Papio cynocephalus anubis clone RP41-470J23, WORKING DRAFT
SEQUENCE, 5 ordered pieces.
AC098645
AC098645.2 GI:18464060
VERSION
HTG: HTGS PHASE2; HTGS_DRAFT.
KEYWORDS
SOURCE
olive baboon.
ORGANISM
Papio cynocephalus anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Papio.
1 (bases 1 to 192695)
REFERENCE
AUTHORS
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, W.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masfello, C., Maskeri, B.,
Mastralian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,
Stratipopoulos, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 192695)
Green, E.D.
Direct Submission
Submitted (27-OCT-2001) NIH Intramural Sequencing Center, 8717
Grovefront Circle, Gaithersburg, MD 20877, USA
On Feb 1, 2002 this sequence version replaced gi:16506407.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nih.gov
----- Project Information
Center project name: csb
Center clone name: 470J23

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

```

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 191421 bases at least Q40
Consensus quality: 191690 bases at least Q30
Consensus quality: 191951 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 192295; sum-of-contigs
Quality coverage: 10.92x in Q20 bases; agarose-fp
Quality coverage: 10.33x in Q20 bases; sum-of-contigs
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

```

* 1 31167: contig of 31167 bp in length
* 31168 31267: gap of unknown length
* 31268 72728: contig of 41461 bp in length
* 72729 72828: gap of unknown length
* 72829 100592: contig of 27764 bp in length
* 100593 100692: gap of unknown length
* 100693 159516: contig of 58824 bp in length
* 159517 159616: gap of unknown length
* 159617 192695: contig of 33079 bp in length.

```

FEATURES

source

1. 192695

/organism="Papio cynocephalus anubis"

/db_xref="taxon:9555"

/clone="RP41-470J23"

/clone_lib="rp41"

misc_feature

1. 31167

/note="assembly_fragment"

clone_end:SP6

vector_side:left

misc_feature

1. 65403

/note="clone overlaps with GenBank Accession Number AC097227 clone RP41-327F23 (center project name csa)"

misc_feature

31268..72728

/note="assembly_fragment"

misc_feature

72829..100592

/note="assembly_fragment"

misc_feature

100693..159516

/note="assembly_fragment"

misc_feature

147682..192695

/note="clone overlaps with GenBank Accession Number AC098698 clone RP41-177A23 (center project name csc)"

misc_feature

159617..192695

/note="assembly_fragment"

misc_feature

clone_end:17

vector_side:right

BASE COUNT

58595 a 40608 c 37486 g 55595 t 411 others

ORIGIN

3.9%; Score 90; DB 2; Length 192695;

Best Local Similarity 62.0%; Pred. No. 5.9e-14;

Matches 160; Conservative 0; Mismatches 95; Indels 3; Gaps 1;

QY 884 ccacatgctggaccaggttgaggcaactctctgttaagacacaggaactattt 943

|| |||| || || |||| || || |||| || || |||| || || ||||

Db 138355 CCGCAGCCTAGAGTAAAGGTGGCAAGTCTCTGTAAAGAGCCAGACAGTGAATATT 138414

QY 944 taggctctgtgacatattgtctcagtcacaactactctctgtctgtgacacga 1003

|| |||| || || |||| || || |||| || || |||| || || ||||

Db 138415 TTGGCTTCTGGGGCCGATCAGCTCTATCACAACCTACTCAACTCTGCCATCTGTAGCACA 138474

QY 1004 aagcaattagcaacaataatgtcaacaacaataatgtgaccccatgaaaaactttatttta 1063

|| |||| || || |||| || || |||| || || |||| || || ||||

Db 138475 AAGCAGTCACACAGCGTATGAAATAATGAGTATGACTGTGTCCATAAAACTTTATT 138534

QY 1064 tggatcaggaacacctgaaaaataatgtctcttttttttttttttttttttttttttttttt 1123

|| |||| || || |||| || || |||| || || |||| || || ||||

Db 138535 ATGGACGTGAAATCTGAGTTGCAATAATTC---TCACATTTTTCACCAACCACTTTAAAAA 138591

QY 1124 cgtaaaaactactcttag 1141

|| |||| || || |||| || || |||| || || |||| || || ||||

Db 138592 TGTAAAGCTATTCTTAG 138609

RESULT 14

AC011060

LOCUS

DEFINITION

AC011060

181009 bp DNA linear HTG 12-MAR-2000

Accession

Accession

Accession

Accession

Accession

Accession

Accession


```

/note="assembly_fragment"
32628..37892
/note="assembly_fragment"
37993..41265
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
41366..52400
/note="assembly_fragment"
52501..63506
/note="assembly_fragment"
63607..79362
/note="assembly_fragment"
clone_end:T7
vector_side:right"
79463..97023
/note="assembly_fragment"
97124..124426
/note="assembly_fragment"
124527..149337
/note="assembly_fragment"
149438..181009
/note="assembly_fragment"
57610 a 36532 c 33173 g 51292 t 2402 others
ORIGIN

Query Match          3.98; Score 89.2; DB 2; Length 181009;
Best Local Similarity 64.68; Pred. No. 9.9e-14;
Matches 133; Conservative 0; Mismatches 73; Indels 0; Caps 0;

Qy 891 ctggaccaggttgggaaactctctctgtaaagaacacagaggaactatttttaggcgc 950
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 151319 CTAGCACAGAGTGCAGAACTTCTCTGTAAAGACCAGATGATAATATTTAGCTTT 151378

Qy 951 tgtgtgcatatgtctctgaacactactctctctgtctctgttagcagcaagaagaat 1010
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 151379 TGTGGCCCATATGTCTCTGTGCACAACTACTCTGCTGTCTTAGGGCAAAAGCAGT 151438

Qy 1011 tagcaacaatgtcaacaacatatgtgacccactgaaactttattttattggtatc 1070
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 151439 TATAGAAATACATATGAATGATGAGCATGTGCTGTGTTCCCATAAAGTATTTATGGACAC 151498

Qy 1071 ggaacctgaaataatgtctttctt 1096
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 151499 TGAATTTGAAATTCATGCTAATTTT 151524

```

```

RESULT 15
AL161628 183483 bp DNA linear PRI 16-JAN-2001
LOCUS Human DNA sequence from clone RP11-31K16 on chromosome 9. Contains
DEFINITION a snRNA binding domain pseudogene, the ELAVL2 gene for ELAV
            (embryonic lethal, abnormal vision, Drosophila)-like 2, ESTs, STSS,
            GSSs and a CpG island, complete sequence.
ACCESSION AL161628
VERSION   AL161628.9 GI:10129841
KEYWORDS  HTG; CpG island; ELAVL2; snRNA binding domain.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 183483)
AUTHORS   Ramsay, H.
TITLE     Direct Submission
JOURNAL   Submitted (20-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
COMMENT   On Sep 14, 2000 this sequence version replaced gi:10039694.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission

```

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., ENBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep/ This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>

This sequence is the entire insert of clone RP11-31K16 The true left end of clone RP11-32112 is at 139671 in this sequence. The true right end of clone RP11-315114 is at 73045 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. RP11-31K16 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

FEATURES

source

```

1..183483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-31K16"
/clone_lib="RPCI-11.1"

```

```

1..224
/note="match: GSS: Em:AQ040596"
misc_feature

1..456
/note="match: GSS: Em:AQ080076"
misc_feature

24..653
/note="match: GSS: Em:AQ045921"
misc_feature

415..471
/note="AluJo repeat: matches 60..118 of consensus"
repeat_region

472..763
/note="AluY repeat: matches 10..288 of consensus"
repeat_region

764..954
/note="AluJo repeat: matches 118..297 of consensus"
repeat_region

1096..1173
/note="MLT1J repeat: matches 103..187 of consensus"
repeat_region

2555..4081
/note="L1MC2 repeat: matches 4721..6314 of consensus"
repeat_region

4184..4223
/note="MLT1I repeat: matches 238..273 of consensus"
repeat_region

4424..4479
/note="MLT1I repeat: matches 343..398 of consensus"
repeat_region

5387..5466
/note="MER58 repeat: matches 2214..2294 of consensus"
repeat_region

5639..5786
/note="L1ME3A repeat: matches 6017..6156 of consensus"
repeat_region

6111..6324
/note="MER20 repeat: matches 1..218 of consensus"
repeat_region

6330..6457
/note="MER5A repeat: matches 2..120 of consensus"
repeat_region

6840..7315
/note="L2 repeat: matches 2252..2736 of consensus"
repeat_region

7345..7434
/note="MIR repeat: matches 2..89 of consensus"
repeat_region

7437..7741
/note="AluX repeat: matches 1..299 of consensus"
repeat_region

7808..7989
/note="MER58 repeat: matches 28..2412 of consensus"
repeat_region

10366..10649
/note="MER58B repeat: matches 46..341 of consensus"
repeat_region

11076..11221
/note="MLT2 repeat: matches 1..147 of consensus"
repeat_region

11222..11661
/note="MLT2CB repeat: matches 1..504 of consensus"
repeat_region

11664..11992

```

```
repeat_region /note="MIR2 repeat: matches 145. .510 of consensus"
misc_feature 12210. .12467
misc_feature /note="L2 repeat: matches 2482. .2750 of consensus"
misc_feature complement(12696. .13167)
misc_feature /note="match: GSS: Em:AQ549073"
misc_feature complement(13792. .14266)
misc_feature /note="match: GSS: Em:AQ594257"
misc_feature complement(14438. .14735)
repeat_region /note="match: GSS: Em:AQ128238"
repeat_region 14756. .14785
repeat_region /note="15 copies 2 mer aa 100% conserved"
repeat_region 16212. .16363
repeat_region /note="FRAM repeat: matches 4. .145 of consensus"
repeat_region 16364. .16385
repeat_region /note="11 copies 2 mer tt 100% conserved"
misc_feature 16462. .16951
repeat_region /note="match: GSS: Em:AQ631149"
repeat_region 16908. .17128
misc_feature /note="MIR repeat: matches 34. .261 of consensus"
misc_feature complement(17116. .17552)
misc_feature /note="match: GSS: Em:AQ133240"
misc_feature 17546. .18050
misc_feature /note="match: GSS: Em:AQ777145"
misc_feature 17660. .17960
repeat_region /note="match: GSS: Em:AQ094704"
repeat_region 19777. .19993
misc_feature /note="MER59A repeat: matches 2. .224 of consensus"
misc_feature 20817. .21173
misc_feature /note="match: GSS: Em:AQ376698"
misc_feature complement(20830. .20939)
misc_feature /note="match: GSS: Em:AQ229933"
misc_feature complement(20830. .21133)
misc_feature /note="match: GSS: Em:AQ235822"
misc_feature complement(20830. .21137)
repeat_region /note="match: GSS: Em:B65379"
repeat_region 21100. .21154
repeat_region /note="11 copies 5 mer aaac 78% conserved"
repeat_region 21183. .21410
repeat_region /note="MER30 repeat: matches 1. .230 of consensus"
repeat_region 21647. .21773
misc_feature /note="MIR repeat: matches 119. .246 of consensus"
repeat_region complement(23139. .23500)
misc_feature /note="match: GSS: Em:AQ210496"
repeat_region 23363. .23601
misc_feature /note="MIR repeat: matches 2. .256 of consensus"
repeat_region complement(24372. .24810)
repeat_region /note="match: GSS: Em:AQ178605"
repeat_region 26871. .27202
repeat_region /note="MSTD repeat: matches 1. .341 of consensus"
repeat_region 27203. .27695
repeat_region /note="L1P10 repeat: matches 5642. .6165 of consensus"
repeat_region 27696. .27750
repeat_region /note="MSTD repeat: matches 341. .394 of consensus"
repeat_region 27867. .27971
repeat_region /note="MERSA repeat: matches 84. .189 of consensus"
repeat_region 27986. .28150
misc_feature /note="MERSA repeat: matches 3. .183 of consensus"
misc_feature 28765. .29329
misc_feature /note="match: GSS: Em:B93338"
misc_feature 28777. .29151
repeat_region /note="match: GSS: Em:AQ125799"
repeat_region 29425. .29482
repeat_region /note="29 copies 2 mer ac 87% conserved"
repeat_region 29740. .29805
repeat_region /note="33 copies 2 mer aa 71% conserved"
repeat_region 29763. .29807
repeat_region /note="9 copies 5 mer aaac 82% conserved"
repeat_region 30205. .30412
repeat_region /note="MIR repeat: matches 2. .230 of consensus"
repeat_region 30582. .30732
misc_feature /note="MERSA repeat: matches 8. .187 of consensus"
misc_feature 30911. .31378
misc_feature /note="match: GSS: Em:AQ546895"
```

```
repeat_region 32009. .32080
gene /note="MIR repeat: matches 132. .206 of consensus"
gene 32125. .32784
CDS /gene="BA31K16.1"
CDS 32125. .32784
CDS /gene="BA31K16.1"
match="BA31K16.1 (snRNA binding domain pseudogene)"
Tr:094534 Sw:021276 Tr:09SG7 Tr:029158 Sw:Q12460
Tr:080401 Tr:065334 Tr:065335 Tr:09U5W4"
/codon_start=1
/pseudo
/evidence=not_experimental
repeat_region 34666. .34977
repeat_region /note="AluSq repeat: matches 11. .297 of consensus"
repeat_region 36947. .36978
misc_feature /note="16 copies 2 mer aa 87% conserved"
misc_feature 38608. .39160
polyA_site /note="match: GSS: Em:AQ353745"
polyA_site complement(40339)
mRNA /gene="ELAVL2"
mRNA complement(join(40339. .43119,43683. .43721,51615. .51840,
55153. .55306,81257. .81360,112241. .112481,171869. .171992))
/gene="ELAVL2"
/note="match: cDNAs: Em:U29088 Em:X85111 Em:M62843
Em:U17602 Em:U12431 Em:L26405 Em:U17599 Em:U29148
Em:U17597 Em:D26158 Em:S83320 Em:D31953 Em:AF176675
match: ESTs: Em:AA545382 Em:AA081584 Em:AI843112
Em:AA656788 Em:AA692890 Em:AI145457 Em:AV117363
Em:AW524152 Em:AA416369 Em:AI028574 Em:AW491988
Em:AA549705 Em:D61361 Em:AV362071 Em:AI837658 Em:AV094529
Em:AA549705 Em:D61361 Em:AV362071 Em:AI837658 Em:AV094529

Query Match 3.9% Score 89.2; DB 9; Length 183483;
Best Local Similarity 64.6%; Pred. No. 9, 9e-14;
Matches 133; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 891 ctggaccagggtgggcaaacctctctctgttaaagaccagacagaaactatttaggttc 950
Db 10360 CTAGCACAGAGGTCAGCAAACTTCTCTGTAAGGACCAGCATGATAAATATTTTAGGTTT 10419

Qy 951 tgtgtgcatatggtcagtcacaaactactcactctctctgtacacacgaaagcaat 1010
Db 10420 TGTGGCCCATATGGTCTCTGTGCACAACTACTCCACTCTGCTGTTCTAGGCAAGCAGT 10479

Qy 1011 tagcaacaatatgtcaacaacacatatgtgaccccatgaaactttttatttatggatcac 1070
Db 10480 TATAGAAAAATACATAAATGAATGAGCATGGCTGTGTTCCCATAAAGATTTTATGGACAC 10539

Qy 1071 ggaacctgaaaaaataatgctcttctt 1096
Db 10540 TCAAAATTTGAATTCATGTAATTTT 10565
```

Search completed: June 6, 2002, 12:16:57
Job time: 14004 sec

FT protein_bind 1109..1113
 FT /*tag= f
 FT /bound_moiety= "C/EBP"
 FT 1207..1214
 FT protein_bind /*tag= g
 FT /bound_moiety= "API"
 FT 1209..1214
 FT /*tag= h
 FT /bound_moiety= "C/EBP"
 FT 1530..1537
 FT /*tag= i
 FT /bound_moiety= "API"
 FT 2115..2121
 FT /*tag= j
 FT /bound_moiety= "API"
 FT 2126..2133
 FT /*tag= k
 FT /bound_moiety= "C/EBP"
 FT 2210..2214
 FT /*tag= l
 FT TATA_signal
 FT W0200164921-A1.
 XX
 PN
 XX
 XX
 PD
 XX
 PF 28-FEB-2001; 2001WO-US06782.
 XX
 XX 29-FEB-2000; 2000US-0515369.
 PR
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 XX Fisher PB, Madireddi MT;
 PI
 XX WPI; 2001-565508/63.
 DR
 XX
 XX
 PT Melanoma differentiation associated gene-7 promoter capable of
 PT treating cancer comprises directing transcription of heterologous
 PT coding sequence encoding tumour suppressor polypeptide positioned
 PT downstream, useful for treating cancer -
 XX
 PS Claim 1; Fig 6A-B; 132pp; English.
 CC
 CC The present sequence is that of the promoter region of the human
 CC melanoma differentiation associated gene-7 (Mda-7). It includes
 CC 2240 bp from the 5' flanking region of the Mda-7 gene. The
 CC promoter was isolated from a human placental genomic library
 CC using a PCR-based method. The Mda-7 promoter exhibits melanocyte
 CC tissue specificity and can only be activated in the targeted
 CC tissue, i.e. the skin. Therefore, a gene of interest driven by
 CC the Mda-7 promoter will be differentially expressed in these cells,
 CC minimising systemic toxicity. A recombinant expression construct
 CC in which the human Mda-7 promoter is operably linked to a coding
 CC sequence encoding a tumour suppressor protein is claimed. The
 CC tumour suppressor is preferably p21, retinoblastoma protein or p53.
 CC A host cell comprising the expression construct is also claimed,
 CC and is preferably a tumour cell selected from a melanoma,
 CC neuroblastoma, astrocytoma, glioblastoma multiforme, cervical
 CC cancer, breast cancer, lung cancer or prostate cancer cell.
 CC The promoter preferably comprises nucleotides 1-2240 of the present
 CC sequence. A pharmaceutical composition including the recombinant
 CC expression construct is used in a claimed method of treating
 CC melanoma, neuroblastoma, astrocytoma, glioblastoma multiforme,
 CC cervical cancer, breast cancer, colon cancer, prostate cancer,
 CC osteosarcoma, chondrosarcoma or a cancer of the central nervous
 CC system.
 XX
 SQ Sequence 2286 BP; 547 A; 592 C; 510 G; 637 T; 0 other;

Query Match 100.08; Score 2286; DB 22; Length 2286;
 Best Local Similarity 100.08; Pred. No. 0;
 Matches 2286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatcagctactatagggcgctgactgcatacacttttgaaccaggctcgcctgct 60
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1 taatcagctactatagggcgctgactgcatacacttttgaaccaggctcgcctgct 60
 QY 61 ccaaagcttgactcataactagattctcaactgatgttgggcccagggttcttagttct 120
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 61 ccaaagcttgactcataactagattctcaactgatgttgggcccagggttcttagttct 120
 QY 121 ctccctgaacctccctctctgaagtataatgctatgataagctcatcgagggtgagcc 180
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 121 ctccctgaacctccctctctgaagtataatgctatgataagctcatcgagggtgagcc 180
 QY 181 aggcacatgttgcctgaactatccatgttatatgcttcttccacacagagtgagct 240
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 181 aggcacatgttgcctgaactatccatgttatatgcttcttccacacagagtgagct 240
 QY 241 actcagatcccagggtgtacccctgaggccagccaaaggtgtatccatgacctcatcct 300
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 241 actcagatcccagggtgtacccctgaggccagccaaaggtgtatccatgacctcatcct 300
 QY 301 gtccagcctgcctttaaacagctcatcccaacctgcctgcctcccgcctcatctgcaga 360
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 301 gtccagcctgcctttaaacagctcatcccaacctgcctgcctcccgcctcatctgcaga 360
 QY 361 cagtagctagatttcagctgcctggggtctattttccctctcagcttctctcttta 420
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 361 cagtagctagatttcagctgcctggggtctattttccctctcagcttctctcttta 420
 QY 421 gctgtctcctgcctcccactcactatttactccagcactctcactcgtctcttctctg 480
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 421 gctgtctcctgcctcccactcactatttactccagcactctcactcgtctcttctctg 480
 QY 481 tctcactcctgctcttgcactcttattctcatagtagtagttaggttcttctggtta 540
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 481 tctcactcctgctcttgcactcttattctcatagtagtagttaggttcttctggtta 540
 QY 541 gccctaaatccacatggtgggaagggggagtgggggaagagtgctgtgaggctgt 600
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 541 gccctaaatccacatggtgggaagggggagtgggggaagagtgctgtgaggctgt 600
 QY 601 gcctacttctgagggttaagactcggccctccagggaacaaaggattcaggctggtgca 660
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 601 gcctacttctgagggttaagactcggccctccagggaacaaaggattcaggctggtgca 660
 QY 661 gctatagccaagcagactgctggcagggtattgcaaggagttatttttctgcttaagaa 720
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 661 gctatagccaagcagactgctggcagggtattgcaaggagttatttttctgcttaagaa 720
 QY 721 aataaacaacactgagtagatgagatggagggggggtgttgggtccagagagattgggaa 780
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 721 aataaacaacactgagtagatgagatggagggggggtgttgggtccagagagattgggaa 780
 QY 781 gagtctgccaagggtgtgttctactcactctcctctcttcttcttctcactcagctgg 840
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 781 gagtctgccaagggtgtgttctactcactctcctctcttcttcttctcactcagctgg 840
 QY 841 aggcagttatcctgtcccccacgttcacattctactccctgttcccatgctgagccacag 900
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 841 aggcagttatcctgtcccccacgttcacattctactccctgttcccatgctgagccacag 900
 QY 901 gttgggcaaacctctctgttaaagaacacagacagaaactatttttagctctgttgccat 960
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 901 gttgggcaaacctctctgttaaagaacacagacagaaactatttttagctctgttgccat 960
 QY 961 atggtctcagtcacaaactactctctctgctctgtagcagaaagcaatagcaacaat 1020
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 961 atggtctcagtcacaaactactctctctgctctgtagcagaaagcaatagcaacaat 1020
 QY 1021 atgtcaacaacatattgtgacccccatgaaactttattttatttgatcggaacactga 1080
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1021 atgtcaacaacatattgtgacccccatgaaactttattttatttgatcggaacactga 1080
 QY 1081 aaataatgtctttctttgtatttttttttttttttttttttttttttttttttt 1140

Db 1081 ||||| 1140
Qy 1141 ggtcgaaggttaagccattcagcttagcagtgaggcgtgatttgctttagacct 1200
Db 1141 ggtcgaaggttaagccattcagcttagcagtgaggcgtgatttgctttagacct 1200
Qy 1201 acagttggccaatccctgattcccaaaatgtattccctcagggatgtggcaaatatttat 1260
Db 1201 acagttggccaatccctgattcccaaaatgtattccctcagggatgtggcaaatatttat 1260
Qy 1261 ggggaagtgtgattaaacagagtttaagaagcatcagaacatttccaggagcggctagcac 1320
Db 1261 ggggaagtgtgattaaacagagtttaagaagcatcagaacatttccaggagcggctagcac 1320
Qy 1321 atgcagaggtctctaaactgacctcaattggatttcattctctggttcagtgaggatcttgaag 1380
Db 1321 atgcagaggtctctaaactgacctcaattggatttcattctggttcagtgaggatcttgaag 1380
Qy 1381 acaagaattctcacaactagactgtgaggactgtgctttgggaacactgctctgttg 1440
Db 1381 acaagaattctcacaactagactgtgaggactgtgctttgggaacactgctctgttg 1440
Qy 1441 atgcctcaactgggcacatggtagaattagagctgagtgcttgccttagctggagatagg 1500
Db 1441 atgcctcaactgggcacatggtagaattagagctgagtgcttgccttagctggagatagg 1500
Qy 1501 gtcagagctcttgactgacctggcagctcttgacacatcacgctgtctgtgtcccttagt 1560
Db 1501 gtcagagctcttgactgacctggcagctcttgacacatcacgctgtctgtgtcccttagt 1560
Qy 1561 ggttcagagccacacagcccaagactagcccaccagagcaccagcctcccagctttctg 1620
Db 1561 ggttcagagccacacagcccaagactagcccaccagagcaccagcctcccagctttctg 1620
Qy 1621 ggttgcctcagctgacatttcttattcttccctggtttccagaaacctaaaggagggcac 1680
Db 1621 ggttgcctcagctgacatttcttattcttccctggtttccagaaacctaaaggagggcac 1680
Qy 1681 atttgggttgagtgattataaccttagggaccatgggtagctgcatgtcaggaacacactc 1740
Db 1681 atttgggttgagtgattataaccttagggaccatgggtagctgcatgtcaggaacacactc 1740
Qy 1741 ctcaacttctggcctgattgattaaaggagagtaacttacaggttatttcttctgtgt 1800
Db 1741 ctcaacttctggcctgattgattaaaggagagtaacttacaggttatttcttctgtgt 1800
Qy 1801 ggactactgtcccagcatgaatagggcattattgaattattttgacaggagggagac 1860
Db 1801 ggactactgtcccagcatgaatagggcattattgaattattttgacaggagggagac 1860
Qy 1861 tgggtatgtgcacagtaataatgtatttacatgtgtacagagtttaccagcacctct 1920
Db 1861 tgggtatgtgcacagtaataatgtatttacatgtgtacagagtttaccagcacctct 1920
Qy 1921 gttgtgttttggcctttgttattacacttgggacaaatttttaaaatttatacatgcag 1980
Db 1921 gttgtgttttggcctttgttattacacttgggacaaatttttaaaatttatacatgcag 1980
Qy 1981 agactgcagcgcagagaagcctaagagactgtccctgtcccacacagccagtggtagagcc 2040
Db 1981 agactgcagcgcagagaagcctaagagactgtccctgtcccacacagccagtggtagagcc 2040
Qy 2041 tgaactcaaacccaggtctcatctcactcaggggctgttctcccacatcgctgtattgtc 2100
Db 2041 tgaactcaaacccaggtctcatctcactcaggggctgttctcccacatcgctgtattgtc 2100
Qy 2101 cttaaagtgtaggtgactagggcaatgaatattctctaggaagcagatgaccaatttcc 2160
Db 2101 cttaaagtgtaggtgactagggcaatgaatattctctaggaagcagatgaccaatttcc 2160
Qy 2161 ctttctccacctccctcttttcttccaccctctcccacatcagcccccatatatatgccc 2220
|||||

Db 2161 ctttctccacctccctcttttcttccacccctcccccattcagcccccatatatatgccc 2220
Qy 2221 aaatctccacaaagccttgcctcgaacaccttacttctgaatgacttccacggct 2280
Db 2221 aaatctccacaaagccttgcctcgaacaccttacttctgaatgacttccacggct 2280
Qy 2281 gggagc 2286
Db 2281 gggagc 2286

RESULT 2
AAK84165/c
ID AAK84165 standard; DNA: 37437 BP.
XX AAK84165;
XX AC
XX DT 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38977.
DE Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.

[illegible]

PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure; SEQ ID NO 38978; 307lpp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients' own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 37442 BP; 10361 A; 6715 C; 6860 G; 13506 T; 0 other;
SQ
Query Match 3.6%; Score 81.4; DB 22; Length 37442;
Best Local Similarity 63.0%; Pred. No. 2.2e-12;
Matches 143; Conservative 0; Mismatches 81; Indels 3; Gaps 1;
QY 809 tctctcttttttctatctccactgagctggaggcaggtatcctgtcccccacgtcaca 868
DB 32329 TCCACTTCATCTTTCATGTCCTACTCTTTCACGATATATTCCTAAAACAACTATAAA 32270
QY 869 ----tctactccggtttcccatgctggaccaggttgggcaaaccttctctgtaaaga 925
DB 32269 ATTGTCCTTTGCAATGGTCAGATATCTAACACAGGGGGTCAGCAAACTTTTCTGAAAGA 32210
QY 926 accagacaggaactatttaggtctgtgtgccatattggctcagtcacaaactactcatc 985
DB 32209 GCCAGATGTAATAATTTTAGGCTTTGTGGCGCAGACATTCCTGTGCAACTATTTCATT 32150
QY 986 tctgctctgtgacgaagaagaattagcaacaatattgtcaacaac 1032
DB 32149 TCTGCTGCTATAGTACAAAGGAGCTTATAAACAAAGTAAGCAAAAC 32103
RESULT 4
AAH14600/c
ID AAH14600 standard; cDNA; 5294 BP.

XX AAH14600;
XX 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:12215.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs.
XX
XX Claim 8; SEQ ID 12215; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to
XX AA895893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 5294 BP; 1637 A; 878 C; 1099 G; 1680 T; 0 other;
SQ
Query Match 3.5%; Score 81; DB 22; Length 5294;
Best Local Similarity 74.5%; Pred. No. 9.5e-13;
Matches 102; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 894 gaccagggttggtggcaaacctctctctgtaagaacacagagaaactattttaggtctgt 953
DB 1915 GACCGAGGAATCAGAAAACTTTTCTGTAAGGGCCAGAGAGTAGTATTAGGTTTGT 1856
QY 954 gtgcacatattggttcacgtcacaaactactctctgtctgtctgtgacgacgaagaattag 1013
ID AAH14600 standard; cDNA; 5294 BP.

Db 1855 GGGCCATATGGTCTCTGTGGAATTACTTAACCTTGCCATCATAGCACGAAGCAGCTAC 1796
Qy 1014 caacaatgtcaacaa 1030
 |||||||
Db 1795 AGACAAATATGACGCA 1779

RESULT 5
AAL36313/c
ID AAL36313 standard; DNA; 26591 BP.
XX AC AAL36313;
XX DT 08-JAN-2002 (first entry)
XX DE Human musculoskeletal system related polynucleotide SEQ ID NO 2678.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX OS Homo sapiens.
XX PN WO200155367-A1.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01338.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229309.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.

DE Human PP1345 protein encoding cDNA SEQ ID NO:19/21.

XX Human; PP1345; cancer suppression; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
FH CDS 174..566
FT /*tag= a
ET /product= "PP1345"

XX CN1313315-A.

XX 19-SEP-2001.

XX 13-MAR-2000; 2000CN-0111989.

XX 13-MAR-2000; 2000CN-0111989.

XX (SHAN-) SHANGHAI INST ONCOLOGY.

XX Gu J, Yang S;

XX WPI: 2002-042193/06.

DR P-PSDB; ABB04713.

XX New human protein able to suppress growth of cancer cells and its
PT encoding polynucleotide sequence -

XX Claim 5; Page 33 (Disclosure); 42pp; Chinese.

XX The present sequence encodes human PP1345 protein, which has cancer-
CC suppressing activity. The present invention also describes a method
CC for the preparation of the protein by recombination, and the application
CC of the protein in treating diseases such as cancer.

XX Sequence 1837 BP; 478 A; 471 C; 407 G; 481 T; 0 other;

Query Match 3.2%; Score 73; DB 24; Length 1837;

Best Local Similarity 64.5%; Pred. No. 1.2e-10;

Matches 109; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 918 tgaagaacacagacagagaactatttttggctctgtgtgacatattgtctcagtcacac 977

Db 1639 tttaaggaacagatagtaataatttttaagcttttgagccacatgtctctgtcacac 1698

QY 978 tactatctctgctgtgacgaaagcaattagcaacaatagtcaacaacatatg 1037

Db 1699 tattcagctctgtattgttaatacaaaagcagctgtgacagtgatgtaaatgaatg 1758

QY 1038 tgaccccatgaaaactttatttatttgatgatacaggaacccctgaaataa 1086

Db 1759 tggctgtgttcataaaactttatttccccaaaaaataa 1807

RESULT 8

AAH98619/c

ID AAH98619 standard; cDNA; 628 BP.

XX AAH98619;

XX 12-OCT-2001 (first entry)

XX Human EST-derived coding sequence SEQ ID NO: 476.

XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

XX diagnostics; forensic test; gene mapping; genetic disorder;

XX biodiversity; gene therapy; nutrition; ss.

XX Homo sapiens.

XX

PN WO200154477-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02687.

XX 25-JAN-2000; 2000US-0491404.

PR 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

XX WPI: 2001-476164/51.

DR P-PSDB; AAM23960.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -

XX Claim 1; Page 514; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.

XX Sequence 628 BP; 190 A; 122 C; 131 G; 185 T; 0 other;

Query Match 3.1%; Score 71.6; DB 22; Length 628;

Best Local Similarity 60.9%; Pred. No. 1.7e-10;

Matches 151; Conservative 0; Mismatches 94; Indels 3; Gaps 2;

QY 894 gaccaggttgaggcaactctctgtaagaacacagaggaactatttagctctgt 953

Db 452 GACCAGGTTTCAGCAGCTTTTCTGTAAAGGCCAAATAGTAATATTTAGCTTTGT 393

QY 954 gtgccatattgtctcagtcacacactatctctgctctgtgacgacgaaagcaatag 1013

Db 392 GGACCATATGCTCTC--TGCTAACTACTCAATTTCTGCCATTGAAGTCAAAAGCAACCAT 335

QY 1014 caacaatgtcacaacacatatgtgaccccatgaaacactttatttatttgatcgga 1073

Db 334 GTA-AATCAATGTACGGCTGTCTTTATTGTGGGACACTGAATTTAAATTTTCATATT 276

QY 1074 aacctgaaaaataatgtctctctcttttatttttcccaacatcaaaaaacgtaaaaact 1133

Db 275 TTCACATAAATAATATTCTTTTCATTTTTCCTCCCAACAAATTAATAATTTTAAAGCC 216

QY 1134 actcttag 1141

Db 215 ATGCTTAG 208

RESULT 9

AAA41353/c

ID AAA41353 standard; cDNA; 286 BP.

XX AAA41353;

XX 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:93.

XX Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;

XX expressed sequence tag; EST; probe; chemotactic; proliferative;

XX


```

FT exon 4050..4403
FT /*tag= g
FT /number= 1
FT CDS 4050..42765
FT /*tag= h
FT /product= "ALDH5A1"
FT /note= "Aldehyde dehydrogenase 5 family member A1.
FT Specifically claimed in claim 25"
FT variation replace(4059,G)
FT /*tag= i
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(4111,A)
FT /*tag= j
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(4155,C)
FT /*tag= k
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(4179,G)
FT /*tag= l
FT /standard_name= "Single nucleotide polymorphism"
FT intron 4404..11575
FT /*tag= m
FT /number= 1
FT variation replace(4430,G)
FT /*tag= n
FT /standard_name= "Single nucleotide polymorphism"
FT exon 11576..11659
FT /*tag= o
FT /number= 2
FT intron 11660..12315
FT /*tag= p
FT /number= 2
FT exon 12316..12486
FT /*tag= q
FT /number= 3
FT variation replace(12415,T)
FT /*tag= r
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(12422,T)
FT /*tag= s
FT /standard_name= "Single nucleotide polymorphism"
FT intron 12487..13921
FT /*tag= t
FT /number= 3
FT variation replace(12559,T)
FT /*tag= u
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(13855,G)
FT /*tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(13917,T)
FT /*tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT exon 13922..14038
FT /*tag= x
FT /number= 4
FT variation replace(14021,T)
FT /*tag= y
FT /standard_name= "Single nucleotide polymorphism"
FT intron 14039..24219
FT /*tag= z
FT /number= 4
FT variation replace(14046,A)
FT /*tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(14179,A)
FT /*tag= ab
FT /standard_name= "Single nucleotide polymorphism"
FT exon 24220..24363
FT /*tag= ac
FT /number= 5
FT intron 24364..29453
FT /*tag= ad

```

```

FT intron /number= 5
FT 24598..31819
FT /*tag= ae
FT /number= 6
FT variation replace(29397,T)
FT /*tag= af
FT /standard_name= "Single nucleotide polymorphism"
FT exon 29454..29597
FT /*tag= ag
FT /number= 6
FT variation replace(29469,A)
FT /*tag= ah
FT /standard_name= "Single nucleotide polymorphism"
FT exon 31820..31978
FT /*tag= ai
FT /number= 7
FT intron 31979..37049
FT /*tag= aj
FT /number= 7
FT exon 37050..37219
FT /*tag= ak
FT /number= 8
FT intron 37220..41171
FT /*tag= al
FT /number= 8
FT exon 41172..41230
FT /*tag= am
FT /number= 9
FT variation replace(41217,C)
FT /*tag= an
FT /standard_name= "Single nucleotide polymorphism"
FT intron 41231..42559
FT /*tag= ao
FT /number= 9
FT variation replace(41237,A)
FT /*tag= ap
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(41403,C)
FT /*tag= aq
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(42380,T)
FT /*tag= ar
FT /standard_name= "Single nucleotide polymorphism"
FT exon 42560..42765
FT /*tag= as
FT /number= 10

```

WO200190119-A2.

29-NOV-2001.

21-MAY-2001; 2001WO-US16558.

19-MAY-2000; 2000US-205849P.

(GENA-) GENAISSANCE PHARM INC.

Kliem SE, Koshy B, Tanguay DA;

WPI; 2002-089912/12.

P-PSDB; AAU73594.

New genetic variants of human aldehyde dehydrogenase 5 family, member A1, ALDH5A1 gene for treating metabolic diseases and for expressing ALDH5A1 protein useful in identifying drugs to treat 4-hydroxybutyric aciduria

Disclosure: Fig 1; 151pp; English.

The invention describes an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the aldehyde dehydrogenase 5 family, member A1 (succinate-semialdehyde dehydrogenase) (ALDH5A1) gene or its fragment.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 08:14:48 ; Search time 2404.01 Seconds
(without alignments)
12834.406 Million cell updates/sec

Title: US-09-515-369B-1
Perfect score: 2286
Sequence: 1 taatcactcactatagg.....tgactccacggtggagc 2286

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred, No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81	3.5	480	10	B1262402
2	80.6	3.5	354	10	T58770
3	80.4	3.5	390	10	L44398
4	78.4	3.4	521	12	A0695956
5	78.2	3.4	435	9	AW183369
6	78.2	3.4	459	9	AI125442
c 7	78.2	3.4	628	12	AQ378792
c 8	77.4	3.4	393	9	AI128823
c 9	77.4	3.4	411	10	W76494
10	77.4	3.4	680	12	AG154552
11	77.2	3.4	400	9	AI926274
c 12	77	3.4	322	9	AI240516
c 13	77	3.4	386	9	AA443938
c 14	77	3.4	393	9	AA444117
c 15	77	3.4	417	9	AI208768
c 16	76.8	3.4	509	10	BI061153
17	76.6	3.4	327	12	B98848

18	76.2	3.3	499	12	AQ277351
c 19	76.2	3.3	644	12	AQ039307
c 20	76.2	3.3	529	12	AF101969
c 21	76	3.3	573	12	AQ633901
22	75.6	3.3	530	12	AQ058112
23	75.6	3.3	628	12	AQ020407
24	75.6	3.3	799	10	BG535645
25	75.2	3.3	488	9	AV604853
c 26	74.6	3.3	348	12	AQ035111
c 27	73.6	3.2	640	12	AG093736
c 28	73.2	3.2	381	9	AW574982
29	73.2	3.2	512	12	AQ768179
30	73	3.2	519	12	AQ029411
c 31	72.8	3.2	302	9	AI559884
32	72.8	3.2	427	12	AQ815705
33	72.2	3.2	312	10	BG230513
34	72	3.1	416	9	AW835797
c 35	72	3.1	764	10	BI915216
c 36	71.8	3.1	2032	11	AF161549
37	71.4	3.1	794	10	BF983586
c 38	71.2	3.1	327	9	AI344682
c 39	71.2	3.1	363	9	AI358986
c 40	71.2	3.1	372	9	AI370636
c 41	71.2	3.1	380	9	AI244892
c 42	71.2	3.1	380	9	AI298789
c 43	71.2	3.1	388	9	AI291781
c 44	71.2	3.1	393	9	AI214952
c 45	71.2	3.1	394	9	AI439755

ALIGNMENTS

RESULT 1

B1262402
LOCUS 602953836T1 NIH_MGC_99 Homo sapiens CDNA clone IMAGE:5088033 3',
DEFINITION mRNA sequence.
ACCESSION B1262402
VERSION B1262402.1 GI:14822582
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 480)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Dr. Louis Staudt, M.D., Ph.D. CDNA Library
Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM1846 row: 9 column: 10
High quality sequence start: 27
High quality sequence stop: 480.
Location/Qualifiers
1. 480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5088033"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dr priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Size-selected >500bp for average insert size

A1125442
 LOCUS qd89b02.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1736619
 3', mRNA sequence.
 ACCESSION A1125442
 VERSION A1125442.1 GI:3593956
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 459)
 AUTHORS NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo,
 Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LENL at:
www-bio.lnlnl.gov/bbrp/image/image.html
 Insert Length: 592 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham.
 FEATURES
 Location/Qualifiers
 1..459
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1736619"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGGAGGCGCCCAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 159 a 88 c 61 g 151 t
 ORIGIN
 Query Match 3.4%; Score 78.2; DB 9; Length 459;
 Best Local Similarity 59.1%; Pred. No. 3.6e-09;
 Matches 204; Conservative 0; Mismatches 113; Indels 28; Gaps 3;
 Qy 901 gttgggcaaacattctctgtgaagaacacagacaggaactatttagctctgtgtgcaat 960
 Db 105 ggtcagcaaaatcttttctgtgaaggccacatacatatatttgcagtcgtgcagacc 164
 Qy 961 atggtctcagtcacactactctctgcctctgtgacgacgaagaacttagcaaat 1020
 Db 165 ttggtctttgtgaaactactcactcgtcgtgaacatacacaagaagctgcacagacaat 224
 Qy 1021 atgtcaacaaacatagtgaacccatgaaaaactttattat---tatggatcacggaacc 1077
 Db 225 atgttaacaaatagggtgactgtgaataaataattgtatttatgaacactgaaatttgaact 284
 Qy 1078 tgaataaatgtctttct-----tttgattttttcccaatcattaaaa 1121
 Db 285 tcaaatcatattttgatgcttcccaaacatttctacttttttgactatttttcaaccattttaa 344
 Qy 1122 aacgtaaaaactactcttagtcgcaaggttaagccattctcagcttagcagtcgacggc 1181
 Db 345 aaagtaaaaccattctttagtgacagcgtctgataaaaact-----agtcattagcc 395

Qy 1182 tggatttggttgacctacagtggccaatccctgattcccaa 1226
 Db 396 ttaattctggcccaagccatagttggccaatgcctgtcacacta 440
 RESULT 7
 LOCUS AQ378792/c
 DEFINITION AQ378792 628 bp DNA linear GSS 20-MAY-1999
 RPCI11-151B24.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-151B24,
 DNA sequence.
 ACCESSION AQ378792
 VERSION AQ378792.1 GI:4349815
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 628)
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
 ,J.C.
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
 Map Building
 JOURNAL Unpublished (1997)
 COMMENT Other_GSSs: RPCI11-151B24.TV
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbestigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from
 Research Genetics (<http://inforesgen.com>). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: SP6
 Class: BAC ends.
 FEATURES
 Location/Qualifiers
 1..628
 /organism="Homo sapiens"
 /db_xref="GDB:7557647"
 /db_xref="taxon:9606"
 /clone="RPCI-11-151B24"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPCI11 Human Male BAC Library"
 BASE COUNT 173 a 133 c 136 g 185 t 1 others
 ORIGIN
 Query Match 3.4%; Score 78.2; DB 12; Length 628;
 Best Local Similarity 72.1%; Pred. No. 4e-09;
 Matches 101; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
 Qy 892 tggaccaggttgggcaaacctctctgttaagaacacagacaggaactatttaggctct 951
 Db 572 TAGCCAGGGGTGGGCATACCTTTTCGTAAAGAGCCAGATAGGAATATTTTCAATTNT 513
 Qy 952 gtgtgcataatgttctcagtcacaaactactctgcctctgtgacgacgaagaact 1011
 Db 512 GTGGCCTTATGGTCTCTCTGTACAAACCACTAAATTTCTGCAGCTGTAGCATGAAGCAGCC 453
 Qy 1012 agcaacaatatgtcaacaa 1031
 Db 452 ACACACAGTAAGTAACAAA 433
 RESULT 8
 A1128823

LOCUS A1128823 393 bp mRNA linear EST 11-SEP-1998
DEFINITION qa94d06.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:1694411 3', mRNA sequence.

ACCESSION A1128823
VERSION A1128823
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 393)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES
Location/Qualifiers
1..393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1694411"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library. Soares fetal lung
NbHH19W."

BASE COUNT 140 a 70 c 53 g 130 t

ORIGIN

Query Match 3.4%; Score 77.4; DB 9; Length 393;
Best Local Similarity 60.2%; Pred. No. 5.6e-09;
Matches 174; Conservative 0; Mismatches 96; Indels 19; Gaps 2;

QY 901 gttgggcaactctctctgtaagaaccagacaggaactatttttagctctgtgtgccat 960
Db 103 GGTACGAAACCTTTTCTGTAAAGGGCCATACATATTTTCAGTCTCTGCAGACCAT 162
QY 961 atggctcagtcacactactctctctgctctgtagcagcaagcaattagcaacaat 1020
Db 163 TTGGTCTTTGCTGAACACTCACTCTGCTAACATACACAAAGCTGCCACAGACAAT 222
QY 1021 atgtcaacaaacatatgtgacccccatgaaactttatttat---tatggatcacggaacc 1077
Db 223 ATGTAAACAAATGAGGGGTGACTGTATAAATGTTATTATGAACACTGAAATTTGAACT 282
QY 1078 tgaaaaataatgtcttctct-----tttgatttttcccccaatcattaaaa 1121
Db 283 TCAATCATTTTTCATGCTTCCCAAAACATTTCTACTTTTGACATATTTTCAACCAATTTAAA 342
QY 1122 aacgtataaacactctcttagctgcgaaggttaagccattctcagcttag 1170
Db 343 AAAGTAAACCAATCTTTAGGTGACAGGCTGTAGAAAAAAGCTAGTCATAG 391

RESULT 9
W76494/c 411 bp mRNA linear EST 17-OCT-1996
LOCUS W76494

DEFINITION 2d67d03.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:345701 5', mRNA sequence.

ACCESSION W76494
VERSION W76494.1
KEYWORDS GI:1386733
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 411)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, F., Soares, M., Tan, F., Trevisakis, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 466 Std Error: 0.00
Seq primer: mob.REGA+ET.

FEATURES
Location/Qualifiers
1..411
/organism="Homo sapiens"
/db_xref="GDB:1271076"
/db_xref="taxon:9606"
/clone="IMAGE:345701"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library. Soares fetal lung
NbHH19W."

BASE COUNT 129 a 58 c 79 g 145 t

ORIGIN

Query Match 3.4%; Score 77.4; DB 10; Length 411;
Best Local Similarity 60.2%; Pred. No. 5.6e-09;
Matches 174; Conservative 0; Mismatches 96; Indels 19; Gaps 2;

QY 901 gttgggcaactctctctgtaagaaccagacaggaactatttttagctctgtgtgccat 960
Db 325 GGTACGAAACCTTTTCTGTAAAGGGCCATACATATTTTCAGTCTCTGCAGACCAT 266
QY 961 atggctcagtcacactactctctctgctctgtagcagcaagcaattagcaacaat 1020
Db 265 TTGGTCTTTGCTGAACACTCACTCTGCTAACATACACAAAGCTGCCACAGACAAT 206
QY 1021 atgtcaacaaacatatgtgacccccatgaaactttatttat---tatggatcacggaacc 1077
Db 205 ATGTAAACAAATGAGGGGTGACTGTATAAATGTTATTATGAACACTGAAATTTGAACT 146
QY 1078 tgaaaaataatgtcttctct-----tttgatttttcccccaatcattaaaa 1121
Db 145 TCAAAATCATTTTTCATGCTTCCCAAAACATTTCTACTTTTGTGACTATTTTTCACCACTTTAAA 86

```

Oy 1122 aacgtataaaactctttaggtcgcaaggttaagccattctcagcttag 1170
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85 AAGCTAAAACCATCTCTTAGGTGACAGGCTGTAGAAAACTAGTCATAG 37

RESULT 10
AG154552
LOCUS
DEFINITION
Pan troglodytes DNA, clone: RP43-018J16.TJ, genomic survey
sequence.
ACCESSION
AG154552
VERSION
AG154552.1 GI:16684230
KEYWORDS
GSS: GSS (genome survey sequence).
SOURCE
Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
Male BAC Library clone:RP43-018J16.TJ.
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library RPCI-43
Unpublished
2 (bases 1 to 680)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel.81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : ECORI
R.Site 2 : ECORI.
Location/Qualifiers
1. .680
/db_xref="taxon:9598"
/clone="RP43-018J16.TJ"
/sex="male"
/cell_type="lymphocytes"
/organism="Pan troglodytes"
BASE COUNT 207 a 127 c 133 g 213 t
ORIGIN

Query Match 3.4%; Score 77.4; DB 12; Length 680;
Best Local Similarity 62.6%; Pred. No. 6.5e-09;
Matches 137; Conservative 0; Mismatches 81; Indels 1; Gaps 1;

Oy 881 ttcccatgctggaccaggttggtggcaactcttcctgtaaaagcagcaggaacta 940
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 227 TCTCCATATTCTAGAAATAGGAGTCAGCAAAATTTCTACTACAAAGGGCAAGACAGTAATA 286

Oy 941 tttaggtctgtgtgccatatggtctcagtcacaaactcactctcctcctgtagca 1000
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 287 TTTTAGGCTTTGAGGCCCATACGGTCTCTGTCATACAACTACTCAATTTGCTGTTTACCG 346

Oy 1001 cgaagcaattagcaacaatagtcaacaacatatgtgaccccatgaaactttatta 1060
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 347 CGAAAGCAGCCATAGACAATACATAAATGAAAAAGCATGGCTGTGTTCCATAACATTTA 406

Oy 1061 -ttatggatcggaacacctgaaataatgtctctttt 1098
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 407 TTTAGGACACTGAATTTGAATTCATATAATTTTAT 445

```

```

RESULT 11
AI926274
LOCUS
DEFINITION
AI926274.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2458138 3',
mRNA sequence.
ACCESSION
AI926274
VERSION
AI926274.1 GI:5662162
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 400)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnli.gov/bbrp/image/image.html
Seq primer: -400P from Gibco
High quality sequence stop: 382.
Location/Qualifiers
1. .400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2458138"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #:
11349-011"
BASE COUNT 133 a 73 c 54 g 140 t
ORIGIN

Query Match 3.4%; Score 77.2; DB 9; Length 400;
Best Local Similarity 64.9%; Pred. No. 6.3e-09;
Matches 131; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

Oy 901 gttggcaaacctctctctgtaaaagcagcaggaactatttttaggtctgtgtgcaat 960
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 127 GGTGAGCAAACTTTCTGTAAAGGGCCATACATACATATTTTCAGTCTCTGCAGACAT 186

Oy 961 atggtctcagtcacaactcactctctgtcgtcgtgacgaaagcaattagcaaat 1020
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 TTGCTCTTTGCTGAACCTACTCTGCTAACTAATACACAAAAAGCTGCACAGACAAT 246

Oy 1021 atgtcaacaacatagtgtaccccatgaaactttattatttgatacggaacatga 1080
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 247 ATGTAACAATAGAGGTGACTGTAAATAAATGTTAT---TTATGAACACTGAAATTTGA 303

Oy 1081 aaataatgtctctcttttatt 1102
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 ACTTCAAAATCATTTTCATGCTT 325

RESULT 12
AI240516/c
LOCUS
DEFINITION
AI240516.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
qh12h02.1, Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:184499 3', mRNA sequence.

```

ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL COMMENT	AI240516 AI240516.1 GI:3835913 EST. human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 322) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. Insert Length: 735 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 315. Location/Qualifiers 1. .322 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="Soares_NFL_T_GBC_S1" /lab_host="DH10B" /note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT ORIGIN	90 a 72 c 67 g 93 t
Query Match Best Local Similarity Matches	3.4%; Score 77; DB 9; Length 322; 71.6%; Pred. No. 6.6e-09; 101; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY	891 ctggaccagggtgggaacctctctgtaagaaacagacagaaactatttagctc 950
Db	178 CTAGACCAAGGTTGGCAACTCTATCTGTATTCCTCCAGATAGTAATAATTTAGGTTT 119
QY	951 tgtgtccatggtctcagtcacaactactcatctctgctctgtagcacgaaagcaat 1010
Db	118 TGTGGCCATAGGGTCTCTATTCCACTACTCACTCTGCTGTGTCAGCATGAAGTAGC 59
QY	1011 tagcaacaatatgtcaacaaa 1031
Db	58 CATAGATGTTAAGTCAACAAA 38
RESULT 13 AA443938 LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS	AA443938 zv51c03.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:757156 3', mRNA sequence. AA443938 AA443938.1 GI:2156613 EST. human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 386) Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
RESULT 14 AA444117/c LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS	AA444117 zv51c03.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:757156 5', mRNA sequence. AA444117 AA444117.1 GI:2156792 EST. human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 393) Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
Query Match Best Local Similarity Matches	3.4%; Score 77; DB 9; Length 386; 63.0%; Pred. No. 7e-09; 119; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAGACAAATACGTCACAAATAAGAGTGGCTGTGTGCCCAATAAACTTT 318
QY	1061 ttatggata 1069
Db	319 ATTCCAAA 327
BASE COUNT ORIGIN	111 a 72 c 68 g 135 t
QY	881 ttctccatgctggaccagggtggcaaacctctctgtaagaaacagacaggaacta 940
Db	139 TTTTTCATCTAGAACATGGGTGGCAAACTTTTCTATAAAGGCGCTCATATAATA 198
QY	941 ttttagcctctgtgtgcacatagtgtcagtcacaactactcatctctgctctgtagca 1000
Db	199 TTTTCAAGCTTTGAGGGTTCACATGCTCTCTGTAGCACCTACTCAATCTGCCATTATAGTA 258
QY	1001 cgaagcaatagcaacaatagtcacaacaacatatgtgaccccatgaaacatttatta 1060
Db	259 CAGAAGCTGCCACACAG

us-09-515-369b-1.rst

Thu Jun 6 16:21:06 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 08:39:08 ; Search time 93.24 Seconds
(without alignments)
6022.289 Million cell updates/sec

Title: US-09-515-369B-1
Perfect score: 2286
Sequence: 1 taatcagctactatagg.....tgacttcacggtggagc 2286

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents,NA:*
1: /cgn2_5/pdata/2/ina/5A_COMB.seq:*
2: /cgn2_5/pdata/2/ina/5B_COMB.seq:*
3: /cgn2_5/pdata/2/ina/6A_COMB.seq:*
4: /cgn2_5/pdata/2/ina/6B_COMB.seq:*
5: /cgn2_5/pdata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_5/pdata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71.8	3.1	7218	1 US-08-232-463-14	Sequence 14, Appl
2	67.6	3.0	6235	4 US-09-305-384-5	Sequence 5, Appl
3	67.6	3.0	6679	4 US-09-305-384-1	Sequence 1, Appl
4	62.2	2.7	178	4 US-08-991-789A-210	Sequence 210, App
5	62.2	2.7	178	4 US-09-062-451-210	Sequence 210, App
6	52.6	2.3	7218	1 US-08-232-463-14	Sequence 14, Appl
7	52	2.3	3786	4 US-08-975-762-42	Sequence 42, Appl
8	52	2.3	3786	4 US-09-295-028-42	Sequence 42, Appl
9	52	2.3	3786	4 US-09-106-582-42	Sequence 42, Appl
10	50.6	2.2	502	2 US-08-967-101-91	Sequence 91, Appl
11	50.6	2.2	502	2 US-08-592-541-91	Sequence 91, Appl
12	50.6	2.2	502	3 US-09-124-698-91	Sequence 91, Appl
13	50.6	2.2	502	4 US-09-127-480-91	Sequence 91, Appl
14	50.6	2.2	502	4 US-08-496-841C-91	Sequence 91, Appl
15	50.6	2.2	1883	2 US-08-967-101-10	Sequence 10, Appl
16	50.6	2.2	1883	2 US-08-967-101-153	Sequence 153, App
17	50.6	2.2	1883	2 US-08-592-541-10	Sequence 10, Appl
18	50.6	2.2	1883	3 US-08-592-541-153	Sequence 153, App
19	50.6	2.2	1883	3 US-08-888-077A-8	Sequence 8, Appl
20	50.6	2.2	1883	3 US-09-124-698-10	Sequence 10, Appl
21	50.6	2.2	1883	3 US-09-124-698-153	Sequence 153, App
22	50.6	2.2	1883	3 US-09-127-480-10	Sequence 10, Appl
23	50.6	2.2	1883	4 US-09-127-480-153	Sequence 153, App
24	50.6	2.2	1883	4 US-08-496-841C-10	Sequence 10, Appl
25	50.6	2.2	1883	4 US-08-496-841C-153	Sequence 153, App
26	48.2	2.1	80246	4 US-09-078-294-4	Sequence 4, Appl
27	48.2	2.1	80595	4 US-09-078-294-3	Sequence 3, Appl

C 28 45.8 2.0 1734 2 US-08-858-052-2 Sequence 2, Appl
C 29 45.8 2.0 1734 3 US-09-200-284-2 Sequence 2, Appl
C 30 43.6 1.9 460 2 US-08-967-101-48 Sequence 48, Appl
C 31 43.6 1.9 460 2 US-08-592-541-48 Sequence 48, Appl
C 32 43.6 1.9 460 3 US-09-124-698-48 Sequence 48, Appl
C 33 43.6 1.9 460 4 US-09-127-480-48 Sequence 48, Appl
C 34 43.6 1.9 460 4 US-08-496-841C-48 Sequence 48, Appl
C 35 42.8 1.9 59065 4 US-09-813-817-3 Sequence 3, Appl
C 36 42.4 1.9 289 4 US-09-007-005-17 Sequence 17, Appl
C 37 42.4 1.9 289 4 US-09-244-796-17 Sequence 17, Appl
C 38 41.4 1.8 703 4 US-09-313-300-6 Sequence 6, Appl
C 39 41.4 1.8 176373 3 US-09-128-155-17 Sequence 17, Appl
C 40 41.2 1.8 7505 4 US-09-078-294-13 Sequence 13, Appl
C 41 39.8 1.7 350 2 US-08-332-766A-20 Sequence 20, Appl
C 42 39.2 1.7 84495 4 US-09-797-906-3 Sequence 3, Appl
C 43 39.2 1.7 152331 3 US-09-128-155-16 Sequence 16, Appl
C 44 38.6 1.7 43795 3 US-08-742-185-101 Sequence 101, App
C 45 38.2 1.7 1732 1 US-08-480-346-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14


```
;
; FILING DATE: 11-Dec-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Potter, Jane E. R.
;   REGISTRATION NUMBER: 33,332
;   REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 622-4900
;   TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 210:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 178 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 210:
US-08-991-789A-210

Query Match      2.7%; Score 62.2; DB 4; Length 178;
Best Local Similarity 71.3%; Pred. No. 9.6e-10;
Matches 82; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 891 ctgagaccaggttggaactctctctgtaagaacacagaggaactatttaggtc 950
   || || || || || || || || || || || || || || || || || || || ||
Db 20 CTAGAGTAGGGTTGGCCAACTTTTCTATAAAGGACGAGAGTAATATTTTCAGGCTT 79
   || || || || || || || || || || || || || || || || || || || ||
QY 951 tgtgtgcattatggtcagtcacaaactctctctctgctctgtagcagaaa 1005
   || || || || || || || || || || || || || || || || || || || ||
Db 80 TGTGGGTTGTGCAGTCTCTCTTGCAACTACTCAGCTCTGCCATTGTAGCATAGAA 134
   || || || || || || || || || || || || || || || || || || || ||

RESULT 5
US-09-062-451-210
; Sequence 210, Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
;   APPLICANT: Fridakis, Tony N.
;   APPLICANT: Smith, John M.
;   TITLE OF INVENTION: REED, STEVEN G.
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;   NUMBER OF SEQUENCES: 297
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SEED AND BERRY LLP
;   STREET: 6300 Columbia Center, 701 Fifth Avenue
;   CITY: Seattle
;   STATE: Washington
;   COUNTRY: USA
;   ZIP: 98104-7092
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   FILING DATE: 04-APR-1997
; CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Maki, David J.
;     REGISTRATION NUMBER: 31,392
;     REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 622-4900
;   TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 210:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 178 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 210:
US-09-062-451-210
```

```
Query Match      2.7%; Score 62.2; DB 4; Length 178;
Best Local Similarity 71.3%; Pred. No. 9.6e-10;
Matches 82; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 891 ctgagaccaggttggaactctctctgtaagaacacagaggaactatttaggtc 950
   || || || || || || || || || || || || || || || || || || || ||
Db 20 CTAGAGTAGGGTTGGCCAACTTTTCTATAAAGGACGAGAGTAATATTTTCAGGCTT 79
   || || || || || || || || || || || || || || || || || || || ||
QY 951 tgtgtgcattatggtcagtcacaaactctctctctgctctgtagcagaaa 1005
   || || || || || || || || || || || || || || || || || || || ||
Db 80 TGTGGGTTGTGCAGTCTCTCTTGCAACTACTCAGCTCTGCCATTGTAGCATAGAA 134
   || || || || || || || || || || || || || || || || || || || ||

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 1800 Diagonal Road, Suite 500
;   CITY: Alexandria
;   STATE: VA
;   COUNTRY: USA
;   ZIP: 22313-0299
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/232,463
;   FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/935,313
;   FILING DATE:
;   APPLICATION NUMBER: EP 91 114 300.6
;   FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
;   NAME: BENT, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703)836-9300
;   TELEFAX: (703)683-4109
;   TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     CLONE: pTZgpt-Fls
;   US-08-232-463-14
```

```
Query Match      2.3%; Score 52.6; DB 1; Length 7218;
Best Local Similarity 9.0%; Pred. No. 9.4e-06;
Matches 28; Conservative 162; Mismatches 121; Indels 0; Gaps 0;

QY 484 catcactgcctcttgacatcttctctcatagtaggttcttcttgtaagcc 543
   || || || || || || || || || || || || || || || || || || || ||
Db 1490 CATCACTGTAATACCTATCTATGCAAGTAGTTAAGAGATAGAGAATTTGGTACRRR 1431
   || || || || || || || || || || || || || || || || || || || ||
```

Qy 544 ctaaatccacatggtgggaaggagggtgggaagagagtgctgtgggctgtgcc 603
Db 1430 RRR 1371
Qy 604 tactctgagggtgaagactcggccctccaggaacaaaggattcagctggtgcagct 663
Db 1370 RRR 1311
Qy 664 atagccaagcagactcgtggccagggtgcaaaaggattgtttgttgaagaaat 723
Db 1310 RRR 1251
Qy 724 aaacacactgatgatgagtgaggagggtgtgtggtgcagagagattgggaagag 783
Db 1250 RRR 1191
Qy 784 tctgccaaggy 794
Db 1190 RRRRRRRRRR 1180

RESULT 7
US-08-975-762-42
; Sequence 42, Application US/08975762
; Patent No. 6207169
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; APPLICATION NUMBER: US/08/975,762
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3786 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-975-762-42

Query Match 2.3%; Score 52; DB 4; Length 3786;
Best Local Similarity 67.2%; Pred. No. 9.7e-06;
Matches 90; Conservative 0; Mismatches 40; Indels 4; Gaps 1;
Qy 914 ttctgttaagaacacagacaggaactatttagctctgtgtgccatattgtc---tca 969
Db 3296 TTTCTGTGTTACTGCCAGACAGATAACATTTTGGACTTTGTATGTATGCTCTTCT 3355
Qy 970 gtcaacactactctctcctctgtagcacgaagcaattagcaacaattatgtcaaca 1029

TREATMENT

Db 3356 GTTCAACTACTGAACCTCTTCCATTGTAGCACGAAGCGGCTGCAGACAATATGTAACA 3415
Qy 1030 aacatatgtgacc 1043
Db 3416 GATGAGCATGACTC 3429

RESULT 8
US-295-028-42
; Sequence 42, Application US/09295028
; Patent No. 6277381
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.439C4
; CURRENT APPLICATION NUMBER: US/09/295,028
; CURRENT FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 42
; LENGTH: 3786
; TYPE: DNA
; ORGANISM: Ehrlichia sp.
; US-09-295-028-42

Query Match 2.3%; Score 52; DB 4; Length 3786;
Best Local Similarity 67.2%; Pred. No. 9.7e-06;
Matches 90; Conservative 0; Mismatches 40; Indels 4; Gaps 1;
Qy 914 ttctgttaagaacacagacaggaactatttagctctgtgtgccatattgtc---tca 969
Db 3296 ttctgtgtactgcgcagacagataaacatttggactttgttatgttctcttct 3355
Qy 970 gtcaacactactctcctctgtagcacgaagcaattagcaacaattatgtcaaca 1029
Db 3356 gttgcaactactgaactcttccattgttagcacgaaggcggtgcagacaattatgtaaca 3415
Qy 1030 aacatatgtgacc 1043
Db 3416 gatgagcatgactc 3429

RESULT 9
US-09-106-582-42
; Sequence 42, Application US/09106582
; Patent No. 6306402
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; STREET: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,582
; FILING DATE: 29-JUN-1998

THERAP

```
;
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J. 31,392
; REGISTRATION NUMBER: 210121.439C2
; REFERENCE/DOCKET NUMBER:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3786 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-106-582-42

Query Match          2.3%; Score 52; DB 4; Length 3786;
Best Local Similarity 67.2%; Pred. No. 9.7e-06;
Matches 90; Conservative 0; Mismatches 40; Indels 4; Gaps 1;

QY 914 ttctgtgaagaaccagacaggaactatttaggctctgtgtgcatatgctc-----tca 969
Db 3296 TTCTGTGTACTGCCAGCAGTAAACATTTGGAGCTTTGTATGTTATATGCTCTCTTCT 3355

QY 970 gtcaaacactactctgctctgtgacgaaagcaattagcaacaatatgtcaaca 1029
Db 3356 GTGCAACTACTGAACTCTTCATTGTAGCAGGAGGCGCTGCAGACAATATGTAACA 3415

QY 1030 aacatagtgtacc 1043
Db 3416 GATGAGCATGACTC 3429

RESULT 10
US-08-967-101-91
; Sequence 91, Application US/08967101
; Patent No. 5840540
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; PRIORITY DATE:
; PRIOR APPLICATION NUMBER: US/08/967,101
; CLASSIFICATION: 435
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-967-101-91

Query Match          2.3%; Score 50.6; DB 2; Length 502;
Best Local Similarity 64.3%; Pred. No. 7.6e-06;
Matches 92; Conservative 0; Mismatches 49; Indels 2; Gaps 1;

QY 917 ctgtaagaaccagacaggaactatttaggctctgtgtgcatatgtctcagtcacaa 976
Db 264 CTGTAAAAAGCCAGACTG--AATATTTTAAAGCTCTATGGGTCTATGGTCTCCAGGGCAA 321
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-967-101-91

Query Match          2.2%; Score 50.6; DB 2; Length 502;
Best Local Similarity 64.3%; Pred. No. 7.6e-06;
Matches 92; Conservative 0; Mismatches 49; Indels 2; Gaps 1;

QY 917 ctgtaagaaccagacaggaactatttaggctctgtgtgcatatgtctcagtcacaa 976
Db 264 CTGTAAAAAGCCAGACTG--AATATTTTAAAGCTCTATGGGTCTATGGTCTCCAGGGCAA 321

QY 977 ctactcatctctgctctgtgacgaaagcaattagcaacaatatgtcaacaacaat 1036
Db 322 ACACCTCACTGTGCTACTGTAGTGTGAAAGCAGGCACAGACAATCTATTACCAAGGAGG 381

QY 1037 gtgacccccatgaaactttattt 1059
Db 382 GTGGTCACTTTCCAAATGAAAGTT 404

RESULT 11
US-08-592-541-91
; Sequence 91, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-592-541-91

Query Match          2.2%; Score 50.6; DB 2; Length 502;
Best Local Similarity 64.3%; Pred. No. 7.6e-06;
Matches 92; Conservative 0; Mismatches 49; Indels 2; Gaps 1;

QY 917 ctgtaagaaccagacaggaactatttaggctctgtgtgcatatgtctcagtcacaa 976
Db 264 CTGTAAAAAGCCAGACTG--AATATTTTAAAGCTCTATGGGTCTATGGTCTCCAGGGCAA 321
```

[illegible]

RESULT 12
US-09-124-698-91
: Sequence 91, Application US/09124698
: Patent No. 6117978
: GENERAL INFORMATION:
: APPLICANT: ST. GEORGE-HYSLOP, PETER H
: APPLICANT: ROMMENS, JOHANNA M
: APPLICANT: FRASER, PAUL E
: TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
: TO ALZHEIMER'S DISEASE
: TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
: NUMBER OF SEQUENCES: 183
: CORRESPONDENCE ADDRESS:
:

Query Match 2.2%; Score 50.6; DB 3; Length 502;
Best Local Similarity 64.3%; Pred. No. 7.6e-06;
Matches 92; Conservative 0; Mismatches 49; Indels 2; Gaps 1;

Qy	977	ctactcattctgcctctgtagcacgaaagaacatttagcaacaatatgtccaacaaacatat	1036
Db	322	ACACTCAACTGTGCTACTTACTGTGAAGCAGGCACAGACAATGCTATTAAACCAAGGAGG	381
Qy	1037	gtgaccccatgaatacttttattt	1059
Db	382	GTGTCACATTCCAATGAAAGTT	404

RESULT 13

```

US-09-127-480-91
; Sequence 91, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:

```

Query Match 2.2%; Score 50.6; DB 4; Length 502;
Best Local Similarity 64.3%; Pred. No. 7.6e-06;
Matches 92; Conservative 0; Mismatches 49; Indels 2; Gaps 1;

Oy	977	ctactcatctgcgcctctgtagcagcaagaagcaatttagcaacaataatgtcaacaacacatat	1036
Db	322	ACACTCAACTGTGCTACTGTAGTGTGAAGAGGACAGACAATGTATTAAACCAAGGAGG	381
Oy	1037	gtgaccccatgaaaaactttattt	1059
Db	382	GTGGTCACTTTTCCAATGAAAGTT	404

```

RESULT 14
US-08-496-841C-91
; Sequence 91, Application US/08496841C
; Patent No. 6210919
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; ROMMENS, JOHANNA M
; FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: Darby & Darby, PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/496,841C

FILING DATE: 28-Jun-1995

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Paul F. Fehlner, Ph.D.

REGISTRATION NUMBER: 35,135

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 527-7700

TELEFAX: (212) 753-6237

INFORMATION FOR SEQ ID NO: 91:

SEQUENCE CHARACTERISTICS:

LENGTH: 502 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 91:

US-08-496-841C-91

Query Match 2.2%; Score 50.6; DB 4; Length 502;

Best Local Similarity 64.3%; Pred. No. 7.6e-06;

Matches 92; Conservative 0; Mismatches 49; Indels 2; Gaps 1;

Qy 917 ctgtaaaagaccagacaggaactattttagctctgtgtgccatattgtctcagtcacaa 976

Db 264 CTGTAAAAAGCCAGACTG--AATATTTTAAAGCTCTATGGTGCATATGCTCCAGGGCAA 321

Qy 977 ctactctctgcctctgtagcagaagaagcaattagcaacaatatgtcaacaacatat 1036

Db 322 ACACCTCAACTGTGCTACTGTAGTGTGAAGCAGGACAGACAATGTATTAAACCAAGGAGG 381

Qy 1037 gtgaccccatgaaaactttattt 1059

Db 382 GTGTCACCTTCCAACTT 404

RESULT 15

US-08-967-101-10/C

Sequence 10, Application US/08967101

Patent No. 5840540

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TITLE OF INVENTION: TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/967,101

FILING DATE: 10-NOV-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/592,541

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1883 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-967-101-10

Query Match 2.2%; Score 50.6; DB 2; Length 1883;

Best Local Similarity 64.3%; Pred. No. 1.7e-05;

Matches 92; Conservative 0; Mismatches 49; Indels 2; Gaps 1;

Qy 917 ctgtaaaagaccagacaggaactattttagctctgtgtgccatattgtctcagtcacaa 976

Db 239 CTGTAAAAAGCCAGACTG--AATATTTTAAAGCTCTATGGTGCATATGCTCCAGGGCAA 182

Qy 977 ctactctctgcctctgtagcagaagaagaattagcaacaatatgtcaacaacatat 1036

Db 181 ACACCTCAACTGTGCTACTGTAGTGTGAAGCAGGACAGACAATGTATTAAACCAAGGAGG 122

Qy 1037 gtgaccccatgaaaactttattt 1059

Db 121 GTGGTCACCTTCCAACTT 99

Search completed: June 6, 2002, 11:55:02

Job time: 11754 sec

